

Equidistribution of joinings under off-diagonal polynomial flows of nilpotent Lie groups

Tim Austin

Abstract

Let G be a connected nilpotent Lie group. Given probability-preserving G -actions $(X_i, \Sigma_i, \mu_i, u_i)$, $i = 0, 1, \dots, k$, and also polynomial maps $\varphi_i : \mathbb{R} \rightarrow G$, $i = 1, \dots, k$, we consider the trajectory of a joining λ of the systems $(X_i, \Sigma_i, \mu_i, u_i)$ under the ‘off-diagonal’ flow

$$(t, (x_0, x_1, x_2, \dots, x_k)) \mapsto (x_0, u_1^{\varphi_1(t)} x_1, u_2^{\varphi_2(t)} x_2, \dots, u_k^{\varphi_k(t)} x_k).$$

It is proved that any joining λ is equidistributed under this flow with respect to some limit joining λ' . This is deduced from the stronger fact of norm convergence for a system of multiple ergodic averages, related to those arising in Furstenberg’s approach to the study of multiple recurrence. It is also shown that the limit joining λ' is invariant under the subgroup of G^{k+1} generated by the image of the off-diagonal flow, in addition to the diagonal subgroup.

Contents

1	Introduction	2
2	Background on group actions	11
3	Real polynomials and Zariski residual sets	13
4	Polynomial maps into nilpotent Lie groups	14
4.1	Polynomial maps	15
4.2	Families of maps and the PET ordering	16
5	Finer results for actions of nilpotent Lie groups	20
6	Idempotent classes	26

7	The case of two-fold joinings	33
8	A partially characteristic factor	34
9	Proof of the main theorem	42
10	Further questions	49
10.1	Other questions in continuous time	49
10.2	Discrete actions	50
A	A continuous-time van der Corput estimate	53

1 Introduction

This paper is set among jointly measurable probability-preserving actions of a connected nilpotent Lie group G . We will assume in addition that G is simply connected; it will be clear from the statements of our main results that by ascending to the universal cover this incurs no real loss of generality.

Suppose that $u_i : G \curvearrowright (X_i, \Sigma_i, \mu_i)$ for $i = 0, 1, \dots, k$ is a tuple of such actions and that λ is a joining of them. This means that λ is a coupling of the measures μ_i on the product space $\prod_i X_i$, and that it is invariant under the **diagonal transformation**

$$u_\Delta^g := u_0^g \times u_1^g \times \dots \times u_k^g$$

for every $g \in G$.

Taking the G -actions on each coordinate separately, the u_i together define a jointly measurable action u_\times of the whole Cartesian power G^{k+1} on $\prod_i X_i$ according to

$$u_\times^{(g_0, g_1, \dots, g_k)} := u_0^{g_0} \times u_1^{g_1} \times \dots \times u_k^{g_k}.$$

In these terms u_Δ may be identified with the restriction of u_\times to the diagonal subgroup

$$G^{\Delta(k+1)} := \{(g, g, \dots, g) : g \in G\} \leq G^{k+1}.$$

An arbitrary joining λ need not be u_\times -invariant. However, the main result of this paper implies that for any one-parameter subgroup $\mathbb{R} \rightarrow G^{k+1}$, the trajectory of λ under the u_\times -action of that subgroup must equidistribute with respect to some new joining λ' that is also invariant under that subgroup. Moreover, this statement

generalizes to averages over the trajectory of any map $\mathbb{R} \longrightarrow G^{k+1}$ that is ‘polynomial’ in the sense that repeated group-valued differencing leads to the trivial map (precise definitions are recalled in Section 4). The full result is the following.

Theorem 1.1 *If $(X_i, \Sigma_i, \mu_i, u_i)$, $0 \leq i \leq k$, and λ are as above, and $\varphi_i : \mathbb{R} \longrightarrow G$ for $1 \leq i \leq k$ are polynomial maps satisfying $\varphi_i(0) = e$ (the identity of G), then the averaged measures*

$$\lambda_T := \int_0^T (\text{id}_{X_0} \times u_1^{\varphi_1(t)} \times \cdots \times u_k^{\varphi_k(t)})_* \lambda \, dt$$

converge in the coupling topology as $T \longrightarrow \infty$ to some joining λ' of the systems $(X_i, \Sigma_i, \mu_i, u_i)$ which is invariant under the restriction of u_\times to the subgroup

$$\langle G^{\Delta(k+1)} \cup \{(e, \varphi_1(t), \dots, \varphi_k(t)) : t \in \mathbb{R}\} \rangle.$$

Here we have used the standard analyst’s notation $\int_a^b := \frac{1}{b-a} \int_a^b$, and we write $\langle S \rangle$ for the smallest closed subgroup of G containing S .

Remark If t is such that $\varphi_i(t) \neq e$ then the individual measures

$$(\text{id}_{X_0} \times u_1^{\varphi_1(t)} \times \cdots \times u_k^{\varphi_k(t)})_* \lambda$$

may *not* be joinings of the original actions. As measures they are still couplings of the μ_i , but the invariance of λ under the diagonal subgroup has been replaced with invariance under its conjugate

$$(e, \varphi_1(t), \dots, \varphi_k(t)) \cdot G^{\Delta(k+1)} \cdot (e, \varphi_1(t), \dots, \varphi_k(t))^{-1}.$$

Thus a non-trivial part of the conclusion of Theorem 1.1 is that the smoothing effect of averaging over t recovers the invariance under $G^{\Delta(k+1)}$ (and likewise under all of these conjugates). \triangleleft

Convergence $\lambda_T \longrightarrow \lambda'$ in the coupling topology, as in Theorem 1.1, asserts that

$$\int_{X_0 \times X_1 \times \cdots \times X_k} f_0 \otimes f_1 \otimes \cdots \otimes f_k \, d\lambda_T \longrightarrow \int_{X_0 \times X_1 \times \cdots \times X_k} f_0 \otimes f_1 \otimes \cdots \otimes f_k \, d\lambda'$$

for any choice of $f_0 \in L^\infty(\mu_0)$, $f_1 \in L^\infty(\mu_1)$, \dots , $f_k \in L^\infty(\mu_k)$. Informally, it is a variant of weak convergence defined against the class of test functions given by tensor products of bounded measurable functions on the individual coordinate-spaces. It is standard that this topology on the convex set of couplings is compact: see, for instance, Theorem 6.2 of Glasner [18].

However, we will actually deduce Theorem 1.1 from a stronger kind of convergence. For any joining λ and any fixed choice of $f_i \in L^\infty(\mu_i)$ for $1 \leq i \leq k$, the map

$$f_0 \mapsto \int_{X_0 \times X_1 \times \cdots \times X_k} f_0 \otimes f_1 \otimes \cdots \otimes f_k \, d\lambda$$

defines a bounded linear functional on $L^2(\mu_0)$, and hence by the self-duality of Hilbert space it specifies a function

$$M^\lambda(f_1, \dots, f_k) \in L^2(\mu_0)$$

(an alternative, more concrete description of M^λ can be found in Section 2 below). The joining convergence asserted by Theorem 1.1 is equivalent to the weak convergence in $L^2(\mu_0)$ of the averages

$$A_T^\lambda(f_1, \dots, f_k) := \int_0^T M^\lambda(f_1 \circ u_1^{\varphi_1(t)}, \dots, f_k \circ u_k^{\varphi_k(t)}) \, dt,$$

but in fact the methods we call on below (particularly the van der Corput estimate, Lemma A.1) naturally give more:

Theorem 1.2 *In the setting of Theorem 1.1, the averages $A_T^\lambda(f_1, \dots, f_k)$ converge in norm in $L^2(\mu_0)$ as $T \rightarrow \infty$ for any functions $f_i \in L^\infty(\mu_i)$, $1 \leq i \leq k$.*

Of course, this does not immediately imply the remainder of Theorem 1.1 concerning the extra symmetries of the limit joining. That will require some additional argument.

The problem of pointwise convergence of the averages A_T^λ remains open, and the methods of the present paper probably say very little about it. One related special case (for certain discrete-time averages) has been established by Bourgain in [12], but I know of no more recent extensions of his work.

Origin and relation to other works

Theorem 1.1 has its origin in the study of multiple recurrence. Furstenberg's original Multiple Recurrence Theorem [15] asserts that for a single probability-preserving transformation $T \curvearrowright (X, \Sigma, \mu)$, if $A \in \Sigma$ has $\mu(A) > 0$ then also

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \cdots \cap T^{-(k-1)n}A) > 0 \quad \forall k \geq 1. \quad (1)$$

In particular, there must be a time $n \geq 1$ at which

$$\mu(A \cap T^{-n}A \cap \dots \cap T^{-(k-1)n}A) > 0 :$$

this is ‘ k -fold multiple recurrence’ for A .

Furstenberg studied this phenomenon in order to give a new proof of a deep theorem of Szemerédi in additive combinatorics [39], which can be deduced quite easily from the Multiple Recurrence Theorem. Following Furstenberg’s original paper, many other works have either proved analogous multiple recurrence assertions in more general settings or analysed the ‘multiple’ ergodic averages of the kind appearing in (1), in particular to determine whether they converge. We will not attempt to give complete references here, but refer the reader to [1], to the paper [21] of Host and Kra and to Chapters 10 and 11 of Tao and Vu’s book [41] for more details.

Many of these convergence questions can be phrased in terms of convergence of joinings, much in the spirit of Theorem 1.1. In Furstenberg’s original setting, if we let μ^Δ be the copy of μ supported on the diagonal in X^k , then the above averages may be re-written as

$$\int_{X^k} 1_A \otimes 1_A \otimes \dots \otimes 1_A \, d\mu_N,$$

where

$$\mu_N := \frac{1}{N} \sum_{n=1}^N (\text{id}_X \times T \times \dots \times T^{k-1})_*^n \mu^\Delta,$$

so in fact the convergence of these scalar averages is almost precisely the assertion that the orbit of the joining μ^Δ under the off-diagonal $\text{id}_X \times T \times \dots \times T^{k-1}$ is equidistributed relative to some limit joining. Convergence here follows from work of Host and Kra [21] (see also Ziegler [44]), and it is worth noting that in this situation the additional invariance of the limit joining under $\text{id}_X \times T \times \dots \times T^{k-1}$ is obvious from the definition of the μ_N and the Følner property of the intervals $\{1, 2, \dots, N\} \subset \mathbb{Z}$.

On the other hand, that additional invariance can be put at the heart of an alternative proof of convergence, which also applies to the more general question of the convergence of the averaged joinings

$$\frac{1}{N} \sum_{n=1}^N (\text{id}_X \times T_1 \times \dots \times T_k)_*^n \mu^\Delta$$

for a commuting tuple of transformations $T_1, T_2, \dots, T_k \curvearrowright (X, \Sigma, \mu)$: see [5, 1] (and compare with Tao [40], where the first proof of convergence for this higher-rank setting was given using very different methods). This more general setting

still exhibits a multiple recurrence phenomenon with striking combinatorial consequences, as shown much earlier by Furstenberg and Katznelson [16]. Another aspect of the study of the limit of the above joinings is that a sufficiently detailed understanding of its structure can be used to give an alternative proof of their theorem [6].

Having come this far, it is natural to ask after the behaviour of these averaged joinings if T_1, T_2, \dots, T_k do not commute, but generate some more complicated discrete group. In particular, if they generate a nilpotent group, then Leibman has shown that multiple recurrence phenomena still occur [23] using an extension of Furstenberg and Katznelson's arguments, but that approach does not prove that the associated functional averages converge in $L^2(\mu)$. The question of convergence seems to be closely related to whether the averages

$$\frac{1}{N} \sum_{n=1}^N (\text{id}_X \times T^{p_1(n)} \times \dots \times T_k^{p_k(n)})_* \mu^\Delta$$

converge for a \mathbb{Z}^d -action T and polynomials $p_i : \mathbb{Z} \rightarrow \mathbb{Z}^d$, at least insofar as some of the standard methods in this area (particularly the van der Corput estimate) run into very similar difficulties in the contexts of these two problems.

These more general convergence questions were posed by Bergelson as Question 9 in [8], having previously been popularized by Furstenberg. Several special cases were established in [17, 9, 20, 25, 4, 13]. On the other hand, the paper [10] contains an example in which $k = 2$, $\langle T_1, T_2 \rangle$ is a two-step solvable group, and convergence fails.

Shortly before the present paper was submitted, Miguel Walsh offered in [43] a proof of convergence for general nilpotent groups and tuples of polynomial maps, so answering the question of Furstenberg and Bergelson in full generality. His proof is most akin to Tao's convergence proof in [40], but clearly involves some non-trivial new ideas as well. It is quite different from the very 'structural' approach taken by most ergodic theoretic papers, such as the present one. It seems likely that Walsh's approach can be adapted to prove convergence in our setting (Theorem 1.2), but it gives much less information on the structure of the resulting factors and joinings (as, for example, in the rest of Theorem 1.3).

Our Theorem 1.2 establishes the analog of the conjecture of Furstenberg and Bergelson (involving both nilpotent groups and polynomial maps) for continuous-time flows. In Subsection 10.2 we will offer some discussion of the additional difficulties presented by an adaptation of our approach to the discrete-time setting. It would still be of interest to find a successful such adaptation, since it would pre-

sumably require uncovering a more detailed description of the relevant factors and joinings, and so would comprise a substantial complement to the approach via Walsh's methods.

We should note also that the case $G = \mathbb{R}^d$ in Theorems 1.1 and 1.2 was recently established in [2]. However, the methods below diverge quite sharply from that previous paper. That work relied crucially on making a time change $t \mapsto t^\alpha$ in the integral averages under study for some small $\alpha > 0$, in order to convert averages along polynomial orbits into averages along orbits given by a linear map perturbed by some terms that grow at sublinear rates in t . That trick leads to a substantial simplification of the necessary induction on families of polynomials (in that paper Bergelson's PET induction is not needed, since something more direct suffices, whereas this induction scheme will appear in the present paper shortly), and so cuts out various other parts of the argument that we use below. However, I do not know how to implement this time-change trick for maps into general nilpotent groups, essentially because various commutators that appear during the proof can give rise to high-degree terms which disrupt the choice of any particular α used to make the leading-order terms linear. It is also my feeling that the argument given below reveals rather more about the relevant structures within probability-preserving G -actions that are responsible for the asymptotic behaviour of the averages in Theorem 1.1.

Although it emerges from the study of multiple recurrence, Theorem 1.1 fits neatly into the general program of equidistribution. Equidistribution phenomena for sequences in compact spaces, and especially sequences arising from dynamical systems, have been popular subjects of analysis for most of the twentieth century: see, for instance, the classic text [22]. Theorem 1.1 can be seen as a close analog of more classical results concerning special classes of compact topological systems: in place of the orbit of an individual point or distinguished subset, we study the orbit of an initially-given joining, and correspondingly vague convergence of measures (that is, tested against continuous functions on a compact space) is replaced by convergence in the coupling topology.

Of course, equidistribution theorems for topological systems always rely very crucially on the special structure of the system under study. Among arbitrary actions on compact spaces there are plentiful examples for which the set of invariant probabilities is very large and unstructured, and which have many points that do not equidistribute. It is interesting that once a tuple of systems $(X_i, \Sigma_i, \mu_i, u_i)$ with invariant probabilities has been fixed, their joinings exhibit the behaviour of Theorem 1.1 without any extra assumptions on those individual systems. Instead, the necessary provisions are that we start with the orbit of some joining, rather than of a

single point, and then prove equidistribution in the sense of the coupling topology.

Among the most profound results giving equidistribution for concrete systems are those concerning the orbits of unipotent flows on homogeneous spaces. In this setting the heart of such an analysis is typically a classification of all invariant probability measures on a system, which then restricts the possible vague limits one can obtain from the empirical measures along an orbit of the system so that, ideally, one can prove that the empirical measures have only one possible limit (and so are equidistributed).

To some extent the approach to Theorem 1.1 parallels that strategy, in that the additional invariances of the limit joinings are an important tool in the proof, and our arguments do imply some further results on the possible structure of the limit joinings (see the second remark following Proposition 8.2).

The full strength of measure classification for probabilities on homogeneous spaces that are invariant and ergodic under the action of a subgroup generated by unipotent elements was finally proved by Ratner in [35, 36], building on several important earlier works of herself and others. The monograph [28] gives a thorough account of this story. Following Ratner's work, Shah proved in [37] some equidistribution results for trajectories of points in homogeneous spaces under flows given by regular algebraic maps into the acting group. That notion of 'polynomial' encompasses ours in many cases, and so his work offers a further point of contact between the two settings.

However, the details of the arguments used below are rather far from those developed by Ratner and her co-workers. For instance, in Shah's paper, he first shows that any vague limit measure for the trajectory of a point under one of his regular algebraic maps must have some invariance under a nontrivial unipotent subgroup. In light of this he can restrict his attention to the possible limit measures that are permitted by Ratner's Measure Classification Theorem, whereupon the extra analysis needed can proceed. By contrast, it is essential in our work that we allow general polynomial maps into G throughout, since our induction would not remain among homomorphisms even if we started there. It would be interesting to know whether an alternative approach to Theorem 1.1 can be found which is more in line with those works on homogeneous space dynamics.

First outline of the proof

Theorems 1.1 and 1.2 will be proved by induction on the tuple of polynomial maps $(\varphi_1, \varphi_2, \dots, \varphi_k)$. The ordering on polynomials that organizes this induction is (a

variant of) Bergelson’s PET ordering from [7], which has become a mainstay of the study of multiple averages involving nilpotent groups or polynomial maps.

To a large extent, the new innovation below is the formulation of an assertion that includes Theorem 1.1 and can be closed on itself in this induction. The delicacy of this formulation is largely attributable to the van der Corput estimate (Lemma A.1), which relates the averages involving a given tuple of polynomial maps to another tuple that precedes it in the PET ordering. In the first place, it is this that forces us to prove Theorem 1.2 alongside Theorem 1.1, but it will also required other features in our inductive hypothesis.

An application of this lemma converts an assertion about a tuple of polynomial maps

$$t \mapsto \varphi_i(t)$$

into another about the ‘differenced’ maps

$$(t, s) \mapsto \varphi_i(t + s)\varphi_i(t)^{-1} \tag{2}$$

(or more complicated relatives of these: see Section 4). Regarded as functions of t alone, these precede the tuple $(\varphi_1, \dots, \varphi_k)$ in the PET ordering for any fixed s . In many applications of PET induction one simply forms these derived maps, then fixes a value of s and applies an inductive hypothesis to the restrictions of these new maps to $\mathbb{R} \times \{s\}$. Unfortunately, in our setting there can be some values of s for which the behaviour of these restrictions is not as ‘good’ as our argument needs. To overcome this we must retain the picture of the new maps in (2) as being polynomial on the whole of $\mathbb{R} \times \mathbb{R}$. As a consequence of this polynomial structure and certain general results about actions of nilpotent Lie groups (see Section 5), one finds that these averages behave ‘asymptotically the same’ for all but a small set of exceptional values of s . This turns out to be a crucial improvement over the possible worst-case behaviour over s . Since repeated appeals to the van der Corput estimate lead to a proliferation of these differencing parameters s , we must actually formulate a theorem which allows for polynomial maps $\mathbb{R} \times \mathbb{R}^r \longrightarrow G$, where we average over the first coordinate in $\mathbb{R} \times \mathbb{R}^r$ and the theorem promises some additional good behaviour for generic values of the remaining r coordinates.

The right notion of genericity to make this precise is provided by Baire’s definition of category, but transplanted into the Zariski topology of \mathbb{R}^n (which is not Hausdorff and so not quite in the usual mould for applications of Baire category). The required notion of ‘Zariski genericity’ will be defined in Section 3, and will be found to relate very well to other standard notions of ‘smallness’ for subsets of \mathbb{R}^n .

In terms of this definition, the complete statement that will be proved by PET

induction is as follows.

Theorem 1.3 *Suppose that $(X_i, \Sigma_i, \mu_i, u_i)$, $0 \leq i \leq k$, and λ are as above and that $\varphi_i : \mathbb{R} \times \mathbb{R}^r \longrightarrow G$, $1 \leq i \leq k$, are polynomial maps satisfying $\varphi_i(0, \cdot) \equiv e$. Let M^λ be constructed from λ as previously, let*

$$A_T^\lambda(f_1, \dots, f_k) := \int_0^T M^\lambda(f_1 \circ u_1^{\varphi_1(t, h)}, \dots, f_k \circ u_k^{\varphi_k(t, h)}) dt$$

(so A_T^λ implicitly depends on h), and let

$$\vec{\varphi} := (e, \varphi_1, \varphi_2, \dots, \varphi_k) : \mathbb{R} \times \mathbb{R}^r \longrightarrow G^{k+1}.$$

Then

1. *for any $h \in \mathbb{R}^r$ and any $f_i \in L^\infty(\mu_i)$, $1 \leq i \leq k$, the functional averages $A_T^\lambda(f_1, \dots, f_k)$ converge in $L^2(\mu_0)$ as $T \longrightarrow \infty$,*
2. *for any $h \in \mathbb{R}^r$ the averaged joinings*

$$\int_0^T (\text{id}_{X_0} \times u_1^{\varphi_1(t, h)} \times u_2^{\varphi_2(t, h)} \times \dots \times u_k^{\varphi_k(t, h)})_* \lambda dt$$

converge as $T \longrightarrow \infty$ to some limit joining λ^h which is invariant under

$$\langle G^{\Delta(k+1)} \cup \text{img } \vec{\varphi}(\cdot, h) \rangle,$$

and

3. *the map $h \mapsto \lambda^h$ is Zariski generically constant on E , and the generic value it takes is a joining invariant under*

$$\langle G^{\Delta(k+1)} \cup \text{img } \vec{\varphi} \rangle.$$

This clearly implies both of the previous theorems. The rest of the paper is directed towards the proof of Theorem 1.3.

Overview of the paper

Sections 2 through 6 establish certain background results that we will need for the main proofs, concerning general properties of group actions and representations; polynomial maps and genericity in the Zariski topology; finer results about actions of nilpotent Lie groups; and the technology of ‘idempotent’ classes of probability-preserving systems. Once all this is at our disposal, the proof of Theorem 1.3 is completed in Sections 7, 8 and 9. Finally Section 10 contains a discussion of various further questions related to those of this paper.

2 Background on group actions

If G is a locally compact second countable (‘l.c.s.c.’) group, then a G -**system** is a quadruple (X, Σ, μ, u) in which (X, Σ, μ) is a standard Borel probability space and $g \mapsto u^g$ is a jointly measurable, μ -preserving left action of G on (X, Σ) . Sometimes this situation will alternatively be denoted by $u : G \curvearrowright (X, \Sigma, \mu)$, and sometimes a whole system will be denoted by a boldface letter such as \mathbf{X} .

Relatedly, a G -**representation** is a strongly continuous orthogonal representation π of G on a separable real Hilbert space \mathfrak{H} . (It would be more conventional to work with complex Hilbert spaces and unitary representations, but choosing the real setting avoids the need to keep track of several complex conjugations later.) This situation will often be denoted by $\pi : G \curvearrowright \mathfrak{H}$. Given a G -system (X, Σ, μ, u) , the associated **Koopman representation** $u^* : G \curvearrowright L^2(\mu)$ is defined by

$$u(g)^* f := f \circ u^{g^{-1}},$$

where this convention concerning inverses ensures that both u and u^* are left actions. Here and throughout the paper the notation L^p , $1 \leq p \leq \infty$, is used for real Lebesgue spaces. It is classical that the joint measurability of u implies the strong continuity of u^* (see, for instance, Lemma 5.28 of Varadarajan [42]), so the Koopman representation is a G -representation in the present sense.

Given a G -system and a closed subgroup $H \leq G$, one may construct the σ -subalgebra

$$\Sigma^H := \{A \in \Sigma : \mu(u^h A \triangle A) = 0 \ \forall h \in H\}.$$

If H is normal in G then this is globally G -invariant, and hence defines a factor of the original system which we call the H -**partially invariant** factor. For some quite special technical reasons we will need only the case of normal H in this paper: see Corollary 5.2 below.

Similarly, for a G -representation π we let

$$\text{Fix}(\pi(H)) := \{v \in \mathfrak{H} : \pi(h)v = v \ \forall h \in H\} \leq \mathfrak{H};$$

for Koopman representations it is easily seen that

$$\text{Fix}(u^*(H)) = L^2(\mu|_{\Sigma^H}).$$

Sometimes it is necessary to compare actions of different groups. If $q : H \longrightarrow G$ is a continuous homomorphism of l.c.s.c. groups and $\mathbf{X} = (X, \Sigma, \mu, u)$ is a G -system, then we may define an H -system on the same probability space by letting h act by $u^{q(h)}$. We denote this system by $\mathbf{X}^{q(\cdot)} = (X, \Sigma, \mu, u^{q(\cdot)})$. A similar construction is clearly possible for representations.

We will also need certain standard calculations involving couplings and joinings. Suppose that λ is a coupling of $\mu_0, \mu_1, \dots, \mu_k$ (without any assumption about group actions). We may regard it instead as a coupling of (X_0, Σ_0, μ_0) with

$$(X_1 \times \dots \times X_k, \Sigma_1 \otimes \dots \otimes \Sigma_k, \lambda')$$

where λ' is the marginal of λ on the last k coordinates. Now λ can be disintegrated over the first coordinate to obtain a probability kernel

$$\Lambda : X_0 \longrightarrow \text{Pr}(X_1 \times \dots \times X_k, \Sigma_1 \otimes \dots \otimes \Sigma_k)$$

so that

$$\lambda = \int_{X_0} \delta_{x_0} \otimes \Lambda(x_0, \cdot) \mu_0(dx_0);$$

and this, in turn, defines a multilinear map

$$M^\lambda : L^\infty(\mu_1) \times \dots \times L^\infty(\mu_k) \longrightarrow L^\infty(\mu_0)$$

according to

$$M^\lambda(f_1, \dots, f_k)(x_0) := \int_{X_1 \times \dots \times X_k} f_1 \otimes f_2 \otimes \dots \otimes f_k d\Lambda(x_0, \cdot).$$

Clearly one has

$$\int_{X_0} f_0 \cdot M^\lambda(f_1, \dots, f_k) d\mu_0 = \int_{X_0 \times X_1 \times \dots \times X_k} f_0 \otimes f_1 \otimes \dots \otimes f_k d\lambda,$$

so this agrees with the definition of M^λ by duality given in the Introduction.

The following is now a routine re-formulation of the definition of a relatively independent product, and the proof is omitted; see, for instance, the third of Examples 6.3 in Glasner [18].

Lemma 2.1 *Let $\Lambda : X_0 \longrightarrow \text{Pr}(X_1 \times \cdots \times X_k)$ be as above and define the relative product measure $\lambda \otimes_0 \lambda$ on $X_1^2 \times \cdots \times X_k^2$ by*

$$\lambda \otimes_0 \lambda = \int_{X_0} \Lambda(x_0, \cdot) \otimes \Lambda(x_0, \cdot) \mu_0(dx_0).$$

Then for any $f_i, g_i \in L^\infty(\mu_i)$, $1 \leq i \leq k$, one has

$$\begin{aligned} \int_{X_0} M^\lambda(f_1, f_2, \dots, f_k) \cdot M^\lambda(g_1, g_2, \dots, g_k) d\mu_0 \\ = \int_{X_1^2 \times \cdots \times X_k^2} f_1 \otimes g_1 \otimes f_2 \otimes g_2 \otimes \cdots \otimes f_k \otimes g_k d(\lambda \otimes_0 \lambda). \end{aligned}$$

□

3 Real polynomials and Zariski residual sets

The third part of Theorem 1.3 involves the notion of Zariski genericity. Recall that on \mathbb{R}^n (or any other real algebraic variety) the **Zariski topology** is the topology whose closed sets are the subvarieties. Although the failure of \mathbb{R} to be algebraically closed gives rise to certain novel behaviour not seen in more classical algebraic geometry (especially under projection maps), in this paper we will not meet any of the situations in which this matters. The basic notions of the theory can be found in many books that use algebraic groups, such as in Subsection D.1 of Starkov [38]. The additional idea we need from that arena is the following.

Definition 3.1 (Zariski meagre and residual sets) *A subset $W \subseteq \mathbb{R}^n$ is **Zariski meagre** if it can be covered by a countable family of proper subvarieties of \mathbb{R}^n . A subset of \mathbb{R}^n is **Zariski residual** if its complement is Zariski meagre. A property that depends on a parameter $h \in \mathbb{R}^n$ is **Zariski generic** if it obtains on a Zariski residual set of h .*

Since proper subvarieties are always closed and nowhere dense in the Euclidean topology, Zariski residual sets are residual in the Euclidean topology. They are therefore ‘large’ in the sense of the Baire Category Theorem and its consequences, but in a much more structured way than an arbitrary Euclidean-residual subset. In particular, they exhibit the following simple behaviour under slicing:

Lemma 3.2 *If $E \subseteq \mathbb{R}^n$ is Zariski meagre and $V \subseteq \mathbb{R}^n$ is any affine subspace then either $E \supseteq V$ or $E \cap V$ is Zariski meagre in V . In the space of translates \mathbb{R}^n/V , the subset of translates for which the former holds is Zariski meagre.*

Proof This is simply a consequence of the corresponding property of Zariski closed sets. \square

Zariski meagre sets are also small in a natural measure-theoretic sense.

Lemma 3.3 *A Zariski meagre subset $E \subseteq \mathbb{R}^n$ has Hausdorff dimension at most $n - 1$.*

Proof Clearly it suffices to show that a single proper algebraic subvariety $V \subseteq \mathbb{R}^n$ has Hausdorff dimension at most $n - 1$, and moreover that this holds when $V = \{f = 0\}$ for some nonzero polynomial $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (because any proper V can be contained in such a zero-set).

This follows by induction on degree. If f is linear then it is immediate, so suppose $\deg f \geq 2$. Then on the one hand the nonsingular locus $\{f = 0\} \cap \{\nabla f \neq 0\}$ can be covered with countably many open sets on which $\{f = 0\} \cap \{\nabla f \neq 0\}$ locally agrees with a smooth $(n-1)$ -dimensional submanifold of \mathbb{R}^n , and hence has Hausdorff dimension $n-1$. On the other hand, the remaining set $\{f = 0\} \cap \{\nabla f = 0\}$ is contained in the set $\{\ell(\nabla f) = 0\}$ for any choice of $\ell \in (\mathbb{R}^n)^* \setminus \{0\}$, which is an algebraic variety generated by a polynomial of degree at most $\deg f - 1$ and so has Hausdorff dimension at most $n - 1$ by the inductive hypothesis. \square

4 Polynomial maps into nilpotent Lie groups

Henceforth G will denote a connected and simply connected nilpotent Lie group, \mathfrak{g} its Lie algebra,

$$G = G^1 \supseteq G^2 \supseteq \cdots \supseteq G^s \supseteq (e)$$

its ascending central series, and

$$\mathfrak{g} = \mathfrak{g}^1 \supseteq \mathfrak{g}^2 \supseteq \cdots \supseteq \mathfrak{g}^s \supseteq (0)$$

the corresponding ascending series of \mathfrak{g} .

In the following we will need certain standard facts about such groups, in particular that the exponential map $\exp : \mathfrak{g} \rightarrow G$ is an analytic diffeomorphism and that any

Lie subalgebra $\mathfrak{h} \leq \mathfrak{g}$ exponentiates to a closed Lie subgroup of G , which is normal if and only if \mathfrak{h} was an ideal. (Note that both of these require the assumption that G is simply connected as well as connected.) These can be found as Theorem 1.2.1 and Corollary 1.2.2 in Corwin and Greenleaf [14], which provides a good general reference for the study of these groups.

4.1 Polynomial maps

Definition 4.1 (Polynomial map) *A map $\varphi : G' \rightarrow G$ between nilpotent Lie groups is **polynomial** if there is some $d \geq 1$ such that*

$$\nabla_{h_1} \nabla_{h_2} \cdots \nabla_{h_d} \varphi \equiv e \quad \forall h_1, h_2, \dots, h_d \in G',$$

where $\nabla_h \varphi(g) := \varphi(gh^{-1})\varphi(g)$.

This definition has come to prominence in the study of multiple recurrence phenomena since Leibman's work generalizing the Furstenberg-Katznelson Multiple Recurrence Theorem to tuples of transformations generating a nilpotent group [23]. For maps into a module M over a ring R (such as an Abelian group, which is a module over \mathbb{Z}), degree- d polynomial maps have been studied much more classically as an ideal of functions $G \rightarrow M$ annihilated under convolution by the d^{th} power of the augmentation ideal of $R[G]$: see, for instance, Passi [29, 30].

In this work we will need the above definition only for $G' = \mathbb{R}^n$. If in addition $G = \mathbb{R}^m$, then it is a simple exercise to show that a map φ is polynomial according to the above if and only if it may be expressed as an m -tuple of polynomials in n variables. For general nilpotent targets G a more concrete view of polynomial maps is still available by the following standard proposition and corollary (for the former see, for instance, Proposition 1.2.7 in Corwin and Greenleaf [14]).

Proposition 4.2 *If G is an s -step connected and simply connected nilpotent Lie group, then $\exp : \mathfrak{g} \rightarrow G$ is a diffeomorphism, and pulled back through \exp the operations of multiplication and inversion become polynomial maps $\mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ and $\mathfrak{g} \rightarrow \mathfrak{g}$ of degree bounded only in terms of s . \square*

Corollary 4.3 *A map $\varphi : \mathbb{R}^n \rightarrow G$ is polynomial if and only if it is of the form $\exp \circ \Phi$ for some polynomial $\Phi : \mathbb{R}^n \rightarrow \mathfrak{g}$.*

Proof This follows by induction on the nilpotency class of G . On the one hand, if $\Phi : \mathbb{R}^n \rightarrow \mathfrak{g}$ is a polynomial, then after $(\deg \Phi)$ -many applications of the differencing operator ∇_\bullet the exponentiated map $\exp \circ \Phi$ may not vanish identically, but at least its projection to G/G^2 vanishes because this is isomorphic to the projection of Φ to $\mathfrak{g}/\mathfrak{g}^2$. Thus finitely many differencing operations yield a polynomial map into \mathfrak{g}^2 , and now repeating this argument s times shows that the differences of $\exp \circ \Phi$ do eventually vanish.

On the other hand, if $\varphi : \mathbb{R}^n \rightarrow G$ is a polynomial map, then the same is true of $\varphi G_2 : \mathbb{R}^n \rightarrow G/G^2 \cong \mathbb{R}^{\dim G - \dim G^2}$. This, in turn, is simply isomorphic to $(\exp^{-1} \circ \varphi) + \mathfrak{g}^2 : \mathbb{R}^n \rightarrow \mathfrak{g}/\mathfrak{g}^2$, so this latter is a polynomial. By choosing lifts of its coefficients under the projection $\mathfrak{g} \rightarrow \mathfrak{g}/\mathfrak{g}^2$, we obtain a polynomial $\Phi_1 : \mathbb{R}^n \rightarrow \mathfrak{g}$ such that $\exp \circ (\exp^{-1} \circ \varphi - \Phi_1)$ takes values in G^2 , and it is clearly still a polynomial map there using the argument of the previous paragraph. Now the inductive hypothesis applied to G^2 gives another polynomial $\Phi_2 : \mathbb{R}^n \rightarrow \mathfrak{g}^2$ such that $\exp \circ (\exp^{-1} \circ \varphi - \Phi_1) = \exp \circ \Phi_2$, and re-arranging this completes the proof. \square

By pulling back to the Lie algebra and arguing there, the above proposition and corollary have the following further consequence, which will be useful in the sequel.

Corollary 4.4 *If $\varphi, \psi : \mathbb{R}^n \rightarrow G$ are polynomial maps, then so are the pointwise product $x \mapsto \varphi(x)\psi(x)$ and the pointwise inverse $x \mapsto \varphi(x)^{-1}$.* \square

4.2 Families of maps and the PET ordering

Our attention now turns to finite tuples

$$\mathcal{F} = (\varphi_1, \varphi_2, \dots, \varphi_k)$$

of polynomial maps $\mathbb{R} \times \mathbb{R}^r \rightarrow G$.

In what follows it is extremely important that we consider the domain of these maps to be split as $\mathbb{R} \times \mathbb{R}^r$. Although this is not really different from \mathbb{R}^{r+1} , the heart of the main induction below rests on comparing the degrees of different polynomial maps into G *in the first coordinate only*. Therefore we will henceforth restrict attention to maps defined on products of \mathbb{R} with other real vector spaces, and will always regard the second coordinate as an auxiliary parameter.

Definition 4.5 (Internal class; leading degree; leading term) *For a polynomial*

map $\varphi : \mathbb{R} \times \mathbb{R}^r \longrightarrow G$ with $\varphi(0, \cdot) \equiv e$, its **internal class** is the greatest c such that $G^c \supseteq \text{img } \varphi$. It is denoted $\text{cl } \varphi$.

Given this, the projection

$$\varphi G^{c+1} : (t, h) \mapsto \varphi(t, h) G^{c+1} : \mathbb{R} \times \mathbb{R}^r \longrightarrow G^c / G^{c+1} \cong \mathbb{R}^{\dim G^c - \dim G^{c+1}}$$

is a Euclidean-valued polynomial map. The **leading degree** $\text{ldeg } \varphi$ of φ is the degree of φG^{c+1} in the variable t , and the **leading term** of φ is the term in φG^{c+1} of the form $t^{\text{ldeg } \varphi} \psi(h)$ for some polynomial map $\psi : \mathbb{R}^r \longrightarrow G^c / G^{c+1}$.

Definition 4.6 (Leading-term equivalence) Two polynomial maps $\varphi, \psi : \mathbb{R} \times \mathbb{R}^r \longrightarrow G$ are **leading-term equivalent**, denoted $\varphi \sim_{\text{LT}} \psi$, if $\text{cl } \varphi = \text{cl } \psi$ and φ and ψ have the same leading term (hence certainly the same leading degree).

Several further definitions are needed in order to explain the PET ordering that will steer the inductive proof of Theorem 1.3. The next roughly follows Leibman [23].

Definition 4.7 (Weight) The **weight** of a polynomial $\varphi : \mathbb{R} \times \mathbb{R}^r \longrightarrow G$ is the pair $\text{wt } \varphi := (\text{cl } \varphi, \text{ldeg } \varphi)$. The set Wt of possible weights (c, d) is ordered lexicographically: pairs $(c, d), (c', d') \in \text{Wt}$ satisfy $(c, d) \prec (c', d')$ if

- either $c > c'$,
- or $c = c'$ and $d < d'$.

Since clearly $\varphi \sim_{\text{LT}} \psi$ implies $\text{wt } \varphi = \text{wt } \psi$, we may also define the **weight** of an \sim_{LT} -equivalence class as the weight of any of its members.

This is a well-ordering on Wt , and it now gives rise to a partial ordering on polynomial maps.

Definition 4.8 (PET ordering on polynomials) Given two polynomial maps $\varphi, \psi : \mathbb{R} \times \mathbb{R}^r \longrightarrow G$, the first **precedes** the second in the **PET ordering**, denoted $\varphi \prec_{\text{PET}} \psi$, if $\text{wt } \varphi \prec \text{wt } \psi$.

Remark Our \prec_{PET} is not quite the same as the PET ordering used in much of the earlier literature for polynomial maps into nilpotent groups. Those required a comparison between polynomials in terms of the individual members of some

Mal'cev basis of G ; see, for instance, Section 3 in [23]. Our ordering is actually a little weaker (in the sense that $\prec_{\text{PET}} \subsetneq \prec_{\text{PET}}^{\text{previous}}$ as relations), because we compare our polynomials on the whole Euclidean subquotients of G arising from the ascending central series, and so in our ordering the assertion that two polynomials have the same leading term is stronger. However, when we later use the PET induction via the van der Corput lemma it will be clear that we are still moving strictly downwards among our families of polynomials, so that the induction proceeds correctly. \triangleleft

The PET ordering on polynomials will play a rôle in the proof of the special case $k = 2$ of Theorem 1.3, but the general case will require an extension of it to an ordering of tuples of polynomials.

Definition 4.9 *Suppose that $f, g : \text{Wt} \rightarrow \mathbb{N}$ are maps which each take nonzero values at only finitely many weights. Then f **precedes** g , denoted $f \prec g$, if there is some $(c, d) \in \text{Wt}$ such that*

- $f(c', d') = g(c', d')$ whenever $(c', d') \succ (c, d)$, and
- $f(c, d) < g(c, d)$.

Definition 4.10 (PET ordering for tuples of polynomials) *If $\mathcal{F} = (\varphi_1, \varphi_2, \dots, \varphi_k)$ is a tuple of polynomial maps then its **weight assignment** is the function $\text{Wt } \mathcal{F} : \text{Wt} \rightarrow \mathbb{N}$ which to each $(c, d) \in \text{Wt}$ assigns the number of \sim_{LT} -equivalence classes of maps in \mathcal{F} that have weight (c, d) .*

*Suppose now that $\mathcal{F} = (\varphi_1, \varphi_2, \dots, \varphi_k)$ and $\mathcal{G} = (\psi_1, \psi_2, \dots, \psi_\ell)$ are families of polynomial maps $\mathbb{R} \times \mathbb{R}^r \rightarrow G$. Then \mathcal{F} **precedes** \mathcal{G} , denoted $\mathcal{F} \prec_{\text{PET}} \mathcal{G}$, if*

- either $\text{Wt } \mathcal{F} \prec \text{Wt } \mathcal{G}$,
- or $\text{Wt } \mathcal{F} = \text{Wt } \mathcal{G}$, and the sets of \sim_{LT} -equivalence classes $\mathcal{F}/\sim_{\text{LT}}$ and $\mathcal{G}/\sim_{\text{LT}}$ can be matched in such a way that (i) their weights match, (ii) every class of \mathcal{F} has cardinality no larger than its corresponding class in \mathcal{G} , and (iii) in at least one instance it is strictly smaller.

As in most proofs that use the PET ordering, it is needed for a particular pair of families of maps, one derived from the other according to the following definitions.

Definition 4.11 (Pivot) *If $\mathcal{F} = (\varphi_1, \varphi_2, \dots, \varphi_k)$ is a tuple of polynomial maps $\mathbb{R} \times \mathbb{R}^r \rightarrow G$ then a **pivot** for \mathcal{F} is a PET-minimal member $\varphi \in \mathcal{F}$.*

Definition 4.12 (Derived family) Suppose that $\mathcal{F} = (\varphi_1, \varphi_2, \dots, \varphi_k)$ is a tuple of polynomial maps $\mathbb{R} \times \mathbb{R}^r \rightarrow G$. Then for $i \leq k$ its i^{th} **derived family** consists of the following polynomial maps $\mathbb{R} \times (\mathbb{R} \times \mathbb{R}^r) \rightarrow G$:

$$(t, k, h) \mapsto \varphi_j(t, h) \varphi_i(t, h)^{-1} \quad \text{for } j \in \{1, 2, \dots, k\} \setminus \{i\}$$

and

$$(t, k, h) \mapsto \varphi_j(k, h)^{-1} \varphi_j(t + k, h) \varphi_i(t, h)^{-1} \quad \text{for } j \in \{1, 2, \dots, k\}.$$

Note that the pre-multiplication by $\varphi_j(k, h)^{-1}$ in the last line has the consequence that if $\varphi_j(0, \cdot) \equiv e$ for every i , then the same is true of the derived family.

Lemma 4.13 If $\mathcal{F} = (\varphi_1, \varphi_2, \dots, \varphi_k)$ with φ_1 a pivot, then its first derived family precedes it in the PET ordering. Also, the sub-tuple $(\varphi_2, \dots, \varphi_k)$ precedes \mathcal{F} in the PET ordering.

Proof For each $j \geq 2$ consider the polynomial maps

$$\varphi_j(t, h) \varphi_1(t, h)^{-1} \quad \text{and} \quad \varphi_j(k, h)^{-1} \varphi_j(t + k, h) \varphi_1(t, h)^{-1}.$$

Because φ_1 is a pivot,

- either $\text{wt } \varphi_j \succ \text{wt } \varphi_1$,
- or $\text{wt } \varphi_j = \text{wt } \varphi_1$ but $\varphi_j \not\sim_{\text{LT}} \varphi_1$,
- or $\varphi_j \sim_{\text{LT}} \varphi_1$.

In the first case both of the new maps above still have weight equal to $\text{wt } \varphi_j$, and are actually leading-term equivalent by comparing their leading terms in G^c/G^{c+1} for $c = \text{cl } \varphi_j$. By the same reasoning, if $\varphi_j \sim_{\text{LT}} \varphi_{j'}$ then all four of the resulting new maps are leading-term equivalent.

The same conclusions hold when $\text{wt } \varphi_1 = \text{wt } \varphi_j$ but $\varphi_1 \not\sim_{\text{LT}} \varphi_j$, since in this case the leading term of either of the above maps into G^c/G^{c+1} is given by the nonzero difference of the leading terms of φ_1 and φ_j .

Lastly, if $\varphi_j \sim_{\text{LT}} \varphi_1$, then these leading terms do cancel, and so both of the polynomial maps written above now strictly precede φ_1 in the PET ordering.

Therefore overall the equivalence classes of \mathcal{F} and of its 1st derived family are in bijective weight-preserving correspondence, apart from the equivalence class of

φ_1 , which is replaced by (possibly several) classes in the derived family of strictly lower weight. This proves the first assertion.

The second assertion is obvious, because the removal of φ_1 either removes a whole \sim_{LT} -equivalence class in case φ_1 is in a singleton class, and hence reduces $\text{Wt } \mathcal{F}$ in \prec , or leaves the \sim_{LT} -class structure of \mathcal{F} unchanged but reduces the cardinality of exactly one of the classes. \square

5 Finer results for actions of nilpotent Lie groups

For any inclusion $H \leq G$ of topological groups, H^n will denote the **topological normal closure** of H in G : that is, the completion of the normal closure in G . This notation suppresses the dependence of this definition on the larger group G , which will always be clear from the context. Similarly, if G is a connected and simply connected Lie group with Lie algebra \mathfrak{g} and $V \leq \mathfrak{g}$ is a Lie subalgebra, then V^n denotes the Lie algebra generated by $\sum_g \text{Ad}(g)V$ (equivalently, the Lie ideal generated by V in \mathfrak{g}), so that $\exp(V^n) = (\exp V)^n$.

The first important result we need is a consequence of the classical Mautner Phenomenon. We will make use of the following expression of this argument as isolated by Margulis [27]; it can also be found as Lemma 2.2 in Subsection 2.1 of Starkov [38].

Lemma 5.1 (Mautner Phenomenon) *Suppose that $\pi : G \curvearrowright \mathfrak{H}$ is a orthogonal representation of a connected Lie group, that $H \leq G$ is a connected Lie subgroup, that $g \in G$ and that there are a sequences $g_i \in G$ and $h_i, h'_i \in H$ with $g_i \rightarrow e$ and $g_i h_i g_i^{-1} h'_i \rightarrow g$. Then*

$$\text{Fix}(\pi(g)) \supseteq \text{Fix}(\pi(H)).$$

\square

Corollary 5.2 *If G is a connected and simply connected nilpotent Lie group, $H \leq G$ is a connected closed subgroup and $\pi : G \curvearrowright \mathfrak{H}$ is an orthogonal representation, then*

$$\text{Fix}(\pi(H)) = \text{Fix}(\pi(H^n)).$$

Similarly, if (X, Σ, μ, u) is a G -system then

$$\Sigma^H = \Sigma^{H^n}.$$

Proof We focus on the first claim, since the second follows at once by considering the Koopman representation.

A simple calculation shows that $H^n = \langle H[H, G] \rangle$, where $[H, G]$ is the subgroup generated by all commutators of elements of H with elements of G . Let

$$G = G_1 \supseteq G_2 \supseteq \dots \supseteq G_s \supseteq G_{s+1} = \{e\}$$

be a central series of G in which each quotient G_r/G_{r+1} has dimension one; for example, one may insert extra terms into the ascending central series, as in the construction of a strong Mal'cev basis. Let \mathfrak{g}_r be the Lie algebra of G_r and \mathfrak{h} the Lie algebra of H .

We will prove by downwards induction on r that if $1 \leq r \leq s$ then

$$\text{Fix}(\pi(\langle H[G_{r+1}, H] \rangle)) = \text{Fix}(\pi(\langle H[G_r, H] \rangle)).$$

When $r = s$ the left-hand side here is $\text{Fix}(\pi(H))$, while when $r = 1$ the right-hand side is $\text{Fix}(\pi(H^n))$, so this will complete the proof.

When $r = s$ the result is clear because G_s is central in G , so now suppose the result is known for some $r + 1 \leq s$. By replacing H with $\langle H[G_{r+1}, H] \rangle$, we may assume that they are equal, since another easy calculation shows that the sets

$$(H[G_{r+1}, H]) \cdot [G_{r+1}, (H[G_{r+1}, H])] \quad \text{and} \quad H[G_{r+1}, H]$$

generate the same subgroup of G .

Let $V \in \mathfrak{g}_r \setminus \mathfrak{g}_{r+1}$, so that \mathfrak{g}_r is the smallest Lie algebra containing both V and \mathfrak{g}_{r+1} . The subgroup $\langle H[G_r, H] \rangle$ is connected, and its Lie algebra is the smallest Lie subalgebra of \mathfrak{g} that contains both \mathfrak{h} and $\{[V, U] : U \in \mathfrak{h}\}$. It therefore suffices to show that any $v \in \text{Fix}(\pi(H))$ is also fixed by $\exp([V, U])$ for any $U \in \mathfrak{h}$.

This can be deduced using Lemma 5.1. We need to show that if $U \in \mathfrak{h}$ then $\exp([V, U])$ is a limit of group elements of the form $g_i h_i g_i^{-1} h'_i$, as treated in that lemma. This follows from the Baker-Campbell-Hausdorff formula, which implies for any $t > 0$ that

$$\exp(tV) \exp((1/t)U) \exp(-tV) \exp(-(1/t)U) = \exp([V, U] + O(t)) \exp(R(t)),$$

where $R(t)$ collects those multiple commutators that involve at least one copy of V and at least two entries from \mathfrak{h} , which must therefore lie in

$$[\mathfrak{g}_{r+1}, \mathfrak{h}] \subseteq \mathfrak{h}.$$

Hence

$$\exp([V, U]) = \exp(tV) \exp((1/t)U) \exp(-tV) \left(\exp(-(1/t)U) \exp(-R(t)) \right),$$

so letting $t = 1/i$ gives the conditions needed by Lemma 5.1. \square

Corollary 5.3 *If G is a connected and simply connected nilpotent Lie group, $H_1, H_2 \leq G$ are connected closed subgroups and $\pi : G \curvearrowright \mathfrak{H}$ is an orthogonal representation, then the subspaces*

$$\text{Fix}(\pi(H_1)), \text{Fix}(\pi(H_2)) \leq \mathfrak{H}$$

are relatively orthogonal over their common further subspace

$$\text{Fix}(\pi(\langle H_1 \cup H_2 \rangle))$$

(meaning that

$$\text{Fix}(\pi(H_1)) \ominus \text{Fix}(\pi(\langle H_1 \cup H_2 \rangle)) \perp \text{Fix}(\pi(H_2)) \ominus \text{Fix}(\pi(\langle H_1 \cup H_2 \rangle)). \quad)$$

Similarly, if (X, Σ, μ, u) is a G -system then Σ^{H_1} and Σ^{H_2} are relatively independent over $\Sigma^{\langle H_1 \cup H_2 \rangle}$.

Proof For a Lie subgroup $H \leq G$, since

$$\text{Fix}(\pi(H)) = \text{Fix}(\pi(H^n))$$

and $H^n \trianglelefteq G$, this subspace of \mathfrak{H} is actually invariant under the whole action π . Therefore the orthogonal projections P_i onto $\text{Fix}(\pi(H_i))$ both commute with π .

It follows that $P_1 P_2$ has image contained in $\text{Fix}(\pi(\langle H_1 \cup H_2 \rangle))$. Since conversely any vector fixed by both H_1 and H_2 is also fixed by P_1 and P_2 , it follows that $P_1 P_2$ is an idempotent with image equal to $\text{Fix}(\pi(\langle H_1 \cup H_2 \rangle))$, and the same holds for $P_2 P_1$. Hence for any vectors $u \in \mathfrak{H}$ and $v \in \text{Fix}(\pi(\langle H_1 \cup H_2 \rangle))$ one has

$$\langle u, v \rangle = \langle u, (P_1 P_2)v \rangle = \langle (P_2 P_1)u, v \rangle,$$

so in fact $P_2 P_1$ is the orthogonal projection onto its image, and similarly for $P_1 P_2$.

Finally, if $v_i \in \text{Fix}(\pi(H_i))$ for $i = 1, 2$ then this implies

$$\langle v_1, v_2 \rangle = \langle P_1 v_1, P_2 v_2 \rangle = \langle P_2 P_1 v_1, v_2 \rangle = \langle (P_2 P_1)v_1, (P_2 P_1)v_2 \rangle,$$

which is the desired relative orthogonality.

In the case of a G -system, applying the above result to the Koopman representation tells us that for any Σ^{H_i} -measurable functions $f_i \in L^2(\mu_i)$ for $i = 1, 2$ we have

$$\int_X f_1 f_2 \, d\mu = \int_X \mathbb{E}(f_1 \mid \Sigma^{\langle H_1 \cup H_2 \rangle}) \mathbb{E}(f_2 \mid \Sigma^{\langle H_1 \cup H_2 \rangle}) \, d\mu,$$

and this is the desired relative independence. \square

Example The above proofs are intimately tied to the nilpotency of G , so it is worth including an example of a solvable Lie group G and representation $\pi : G \curvearrowright \mathfrak{H}$ to show that this restriction is really needed.

Let $\rho : \mathbb{R} \curvearrowright \mathbb{C}$ be the rotation action defined by

$$\rho^t z := e^{2\pi i t} z$$

and let $G := \mathbb{C} \rtimes_\rho \mathbb{R}$. This is a simple three-dimensional solvable Lie group; in coordinates it is $\mathbb{C} \times \mathbb{R}$ with the product

$$(u, s) \cdot (v, t) := (\rho^t u + v, s + t).$$

It may also be interpreted as a group extension of \mathbb{Z} by the group $\mathbb{C} \rtimes S^1$ of orientation-preserving isometries of \mathbb{C} , and this picture gives an action $\xi : G \curvearrowright \mathbb{C}$ with kernel isomorphic to \mathbb{Z} .

For each $v \in \mathbb{C}$ let G_v be the isotropy subgroup $\{g \in G : \xi^g v = v\}$. Then $G_v \cong \mathbb{R}$, and G_v and G_w are conjugated by the ‘translational’ element $(w - v, 0) \in G$. Moreover, since any translation of \mathbb{C} may be obtained as a composite of two rotations about different points, the groups G_v together generate G , and so $G_v^n = G$ for every v . A simple calculation shows that in coordinates one has

$$G_v = \{(v - \rho^t(v), t) : t \in \mathbb{R}\}.$$

Now consider the action $\pi : G \curvearrowright L^2_{\mathbb{C}}(m_{S^1}) \cong L^2(m_{S^1}) \otimes_{\mathbb{R}} \mathbb{C}$ defined by

$$(\pi(u, t)f)(z) := e^{2\pi i \langle \rho^{-t} u, z \rangle} f(\rho^t z),$$

where $\langle \rho^{-t} u, z \rangle$ is the usual inner product of \mathbb{C} regarded as a vector space over \mathbb{R} . (A routine check shows that this formula correctly defines an action of G .) The subspace $\text{Fix}(\pi(G_v))$ consists of those functions f such that

$$e^{2\pi i \langle \rho^{-t} u, z \rangle} f(\rho^t z) = f(z) \quad \forall z \in S^1, t \in \mathbb{R} :$$

that is, of the constant complex multiples of the function $z \mapsto e^{-2\pi i \langle u, z \rangle}$. These are all distinct 2-real-dimensional subspaces of $L^2_{\mathbb{C}}(m_{S^1})$, so are not equal to $\text{Fix}(\pi(G)) =$

$\{0\}$, and also (by considering close-by values of v , for instance) they are not pairwise orthogonal. \triangleleft

Another useful result in a similar vein to Corollary 5.2 is the following simple relative of the Pugh-Shub Theorem [32]. An adaptation of their theorem to the setting of nilpotent groups has previously been given by Ratner in Proposition 5.1 of [35]. Although our formulation is superficially different from hers, each version can easily be deduced from the proof of the other.

Lemma 5.4 *Let $\pi : G \curvearrowright \mathfrak{H}$ be an orthogonal representation of a connected nilpotent Lie group, and let $\text{Lat } \mathfrak{g}$ be the family of all proper Lie subalgebras of \mathfrak{g} . Then the subfamily*

$$\mathcal{A} := \{V \in \text{Lat } \mathfrak{g} : \text{Fix}(\pi(\exp V)) \supsetneq \text{Fix}(\pi(G))\}$$

has countably many maximal elements.

Proof Suppose that $V_1, V_2 \in \mathcal{A}$ are two distinct maximal elements. Then the Lie subalgebra generated by $V_1 + V_2$ must strictly contain them both, and hence

$$\text{Fix}(\pi(\langle \exp V_1 \cup \exp V_2 \rangle)) = \text{Fix}(\pi(G)),$$

by their maximality.

Corollary 5.3 now implies that $\text{Fix}(\pi(\exp V_1))$ and $\text{Fix}(\pi(\exp V_2))$ are relatively orthogonal over $\text{Fix}(\pi(G))$. Therefore there can be at most countably many of these maximal elements of \mathcal{A} , because \mathfrak{H} is separable: indeed, if $\mathcal{A}_1 \subseteq \mathcal{A}$ were an uncountable collection of maximal elements, then choosing some representative unit vectors

$$x_V \in \text{Fix}(\pi(\exp V)) \ominus \text{Fix}(\pi(G)) \quad \forall V \in \mathcal{A}_1$$

would give an uncountable sequence of orthonormal vectors in \mathfrak{H} , and hence a contradiction. \square

Example It is certainly not true that \mathcal{A}_1 is generally finite. For example, consider the obvious rotation action of \mathbb{R}^2 on \mathbb{T}^2 and let $\pi : \mathbb{R}^2 \curvearrowright L^2(m_{\mathbb{T}^2})$ be the resulting orthogonal representation. Then any one-dimensional subgroup $\mathbb{R}\mathbf{v} \leq \mathbb{R}^2$ of rational slope has some non-trivial invariant functions, but the whole \mathbb{R}^2 -action is ergodic. \triangleleft

This conclusion of countability (rather than finitude) gives rise to the need for the notion of Zariski genericity (rather than simply Zariski openness). The connection between them is established by the following.

Corollary 5.5 *If $\varphi : \mathbb{R} \times \mathbb{R}^r \longrightarrow G$ is a polynomial map into a connected and simply connected nilpotent Lie group and $\pi : G \curvearrowright \mathfrak{H}$ is an orthogonal representation, then the map*

$$\mathbb{R}^r \longrightarrow (\text{subspaces of } \mathfrak{H}) : h \mapsto \text{Fix}(\pi(\langle \text{img } \varphi(\cdot, h) \rangle))$$

takes the fixed value $\text{Fix}(\pi(\langle \text{img } \varphi \rangle))$ Zariski generically. Similarly, if (X, Σ, μ, u) is a G -system then the σ -subalgebra $\Sigma^{\langle \text{img } \varphi(\cdot, h) \rangle}$ agrees with $\Sigma^{\langle \text{img } \varphi \rangle}$ up to μ -negligible sets for Zariski generic h .

Proof Replacing G with $\langle \text{img } \varphi \rangle^n$ if necessary, we may assume they are equal.

Let $\mathcal{A} \leq \text{Lat } \mathfrak{g}$ be the family of all Lie subalgebras with fixed-point subspaces strictly larger than $\text{Fix}(\pi(G))$, as in Lemma 5.4, and let $\mathcal{A}_1 \subseteq \mathcal{A}$ be the subfamily of maximal elements of \mathcal{A} , so Lemma 5.4 shows that this is countable. Since $\text{Fix}(\pi(\exp V^n)) = \text{Fix}(\pi(\exp V))$ for any $V \in \text{Lat } \mathfrak{g}$ by Corollary 5.2, by maximality we must have $V = V^n$ for every $V \in \mathcal{A}_1$.

Now,

$$\begin{aligned} \{h : \text{Fix}(\pi(\langle \text{img } \varphi(\cdot, h) \rangle)) \supsetneq \text{Fix}(\pi(\langle \text{img } \varphi \rangle)) = \text{Fix}(\pi(G))\} \\ = \bigcup_{V \in \mathcal{A}_1} \{h : \varphi(t, h) \in \exp V \ \forall t \in \mathbb{R}\}, \end{aligned}$$

and so by the countability of \mathcal{A}_1 it suffices to show that each individual set $\{h : \exp \varphi(t, h) \in V \ \forall t \in \mathbb{R}\}$ is proper and Zariski closed in \mathbb{R}^r . Since $\text{Fix}(\pi(\langle \text{img } \varphi \rangle)) = \text{Fix}(\pi(G))$, the subgroup $\langle \text{img } \varphi \rangle$ is not contained in $\exp V$ for any $V \in \mathcal{A}_1$, and so in fact $\text{img } \varphi \not\subseteq \exp V$ (since $\exp V$ is itself a subgroup).

Therefore for any $V \in \mathcal{A}_1$ we may choose a linear form $\ell \in \mathfrak{g}^*$ which annihilates V but does not annihilate the whole of $\exp^{-1}\langle \text{img } \varphi \rangle$, and now one has

$$\{h : \varphi(t, h) \in \exp V \ \forall t \in \mathbb{R}\} \subseteq \{h : \ell(\exp^{-1}(\varphi(t, h))) = 0 \ \forall t \in \mathbb{R}\}.$$

However, the map $(t, h) \mapsto \ell(\exp^{-1}(\varphi(t, h)))$ is a polynomial $\mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}$, by Corollary 4.3. By collecting monomials it may be expressed as

$$t^d p_d(h) + t^{d-1} p_{d-1}(h) + \cdots + t p_1(h) + p_0(h)$$

for some $p_i \in \mathbb{R}[h_1, \dots, h_r]$, and now

$$\{h : \ell(\exp^{-1}(\varphi(t, h))) = 0 \ \forall t \in \mathbb{R}\} = \bigcap_{i=0}^d \{h : p_i(h) = 0\}.$$

This is manifestly a real algebraic subvariety of \mathbb{R}^n , and it is proper because the map $\ell \circ \exp^{-1} \circ \varphi$ was chosen so as not to vanish identically, so it is a Zariski meagre subset of \mathbb{R}^n , as required.

Once again the conclusion about G -systems follows at once by considering Koopman representations. \square

6 Idempotent classes

The final ingredients needed for the proof of Theorem 1.3 are some results on ‘idempotent classes’ of probability-preserving systems. These were introduced in [3, 4] building on the earlier notion of a ‘pleasant extensions’ of systems [5] (and also worth comparing with Host’s ‘magic extensions’ from [19]).

Definition 6.1 (Idempotent and hereditary classes) *For any l.c.s.c. group G , a class \mathcal{C} of jointly-measurable, probability-preserving G -systems is **idempotent** if it is closed under measure-theoretic isomorphisms, inverse limits and arbitrary joinings. It is **hereditary** if it is closed under passing to factors.*

Example The leading examples of idempotent classes are those of the form

$$\mathcal{C}_0^{H_1} \vee \dots \vee \mathcal{C}_0^{H_\ell}$$

for some closed normal subgroups $H_1, H_2, \dots, H_\ell \trianglelefteq G$, where this denotes the class of all G -systems which can be expressed as a joining of systems $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_\ell$ where each \mathbf{Y}_i has trivial H_i -subaction. \triangleleft

The reference [1] contains an introduction to idempotent classes in the case of a discrete acting group. In earlier works, idempotent classes were introduced to set up the theory of ‘sated extensions’ of probability-preserving systems, which then play the primary rôle in applications of these ideas. However, sated extensions are a little inconvenient in the present setting, and so we will work instead with some more elementary results about idempotent classes. The reasoning behind this change of perspective relates to the need to change the group that acts on a system, which will appear in Section 8.

In addition, our interest here is in actions of Lie groups, for which these ideas have not previously appeared in the literature. Therefore the basic definitions and results we need have been included below for completeness. Only very simple changes and additions are needed to the treatments in [1] or [3]. We will also introduce a

slightly novel example of an idempotent class, useful for handling the polynomial maps of the present setting.

Lemma 6.2 (C.f. Lemma 2.2.2 in [1]) *If \mathcal{C} is an idempotent class of G -systems and $\mathbf{X} = (X, \Sigma, \mu, u)$ is any G -system, then \mathbf{X} has an essentially unique largest factor $\Lambda \leq \Sigma$ that may be generated by a factor map to a member of \mathcal{C} .*

Proof It is clear that under the above assumption the family of factors

$$\{\Xi \leq \Sigma : \Xi \text{ is generated by a factor map to a system in } \mathcal{C}\}$$

is nonempty (it contains $\{\emptyset, X\}$, which corresponds to the trivial system), upwards directed (because \mathcal{C} is closed under joinings) and closed under taking σ -algebra completions of increasing unions (because \mathcal{C} is closed under inverse limits). There is therefore a maximal σ -subalgebra in this family. \square

Definition 6.3 (Maximal \mathcal{C} -factors) *The factor Λ obtained in the preceding lemma is the **maximal \mathcal{C} -factor** of (X, Σ, μ, u) , and will sometimes be denoted by the (slightly abusive) notation $\mathcal{C}\Sigma$. Similarly, we will sometimes denote by $\mathbf{C}\mathbf{X}$ a choice of a member of \mathcal{C} such that $\mathcal{C}\Sigma$ can be generated by a factor map $\mathbf{X} \rightarrow \mathbf{C}\mathbf{X}$.*

The importance of idempotent classes derives from the following proposition.

Proposition 6.4 (Joinings to members of idempotent classes) *Suppose that \mathcal{C} is a hereditary idempotent class of G -systems, that $\mathbf{X} = (X, \Sigma, \mu, u)$ is any G -system and $\mathbf{Y} = (Y, \Phi, \nu, v)$ is a member of \mathcal{C} . Then for any joining*

$$\begin{array}{ccc} & \mathbf{Z} = (X \times Y, \Sigma \otimes \Phi, \lambda, u \times v) & \\ \pi \swarrow & & \searrow \xi \\ \mathbf{X} & & \mathbf{Y}, \end{array}$$

where π and ξ are the coordinate projections, there is some further factor Λ of \mathbf{X} which is generated by a factor map to a member of \mathcal{C} and such that the factor $\pi^{-1}(\Sigma)$ is relatively independent from $\xi^{-1}(\Phi)$ over $\pi^{-1}(\Lambda)$. Concretely, this means that

$$\int_Z f(x)g(y) \lambda(dx, dy) = \int_Z \mathbb{E}_\mu(f | \Lambda)(x)g(y) \lambda(dx, dy)$$

for any $f \in L^\infty(\mu)$ and $g \in L^\infty(\nu)$ (so we do not require that $\pi^{-1}(\Lambda)$ also be contained in $\xi^{-1}(\Phi)$ up to negligible sets).

Proof We will construct from the joining λ a new joining of \mathbf{X} with a C-system such that λ is relatively independent over a factor of \mathbf{X} which in that new joining is actually determined by the coordinate in the C-system.

Let $\Lambda : X \longrightarrow \text{Pr } Y$ be a disintegration of λ over the coordinate projection to X . Form the infinite Cartesian product

$$Z' := X \times Y^{\mathbb{N}}$$

and let λ' be the $(u \times v^{\times \mathbb{N}})$ -invariant measure obtained as the relatively independent product of copies of λ :

$$\lambda' = \int_X \delta_x \otimes \Lambda(x, \cdot)^{\otimes \mathbb{N}} \mu(dx).$$

Let $\pi' : Z' \longrightarrow X$ be the first coordinate projection, and let λ_1 be the image of λ' under the projection to $Y^{\mathbb{N}}$.

Finally, let $\Lambda \leq \Sigma$ be the σ -algebra of those sets which are λ' -a.s. determined by the remaining coordinates of Z' :

$$\Lambda := \{A \in \Sigma : \exists B \in \Phi^{\otimes \mathbb{N}} \text{ s.t. } \lambda'((A \times Y^{\mathbb{N}}) \triangle (X \times B)) = 0\}.$$

This is clearly a factor of \mathbf{X} , and by definition it also specifies a factor of the system $(Y^{\mathbb{N}}, \Phi^{\otimes \mathbb{N}}, \lambda_1, v^{\times \mathbb{N}})$ (since each $A \in \Lambda$ is identified with a member of $\Phi^{\otimes \mathbb{N}}$, uniquely up to negligible sets). Let $\Lambda' := (\pi')^{-1}(\Lambda)$, so up to negligible sets this is measurable with respect to either π' or the coordinate projection $Z' \longrightarrow Y^{\mathbb{N}}$. The system $(Y^{\mathbb{N}}, \Phi^{\otimes \mathbb{N}}, \lambda_1, v^{\times \mathbb{N}})$ is a member of \mathbf{C} , because $\mathbf{Y} \in \mathbf{C}$ and \mathbf{C} is closed under joinings; and hence the factor of \mathbf{X} generated by Λ is also in \mathbf{C} , because it may be identified with a factor of that member of \mathbf{C} and \mathbf{C} is hereditary.

Now let $f \in L^\infty(\mu)$ and $g \in L^\infty(\nu)$. To prove the desired equality of integrals, it suffices to show that

$$E_\mu(f | \Lambda) = 0 \implies E_\lambda(f \circ \pi | \{\emptyset, X\} \otimes \Phi) = 0,$$

since an arbitrary f may be decomposed as $E_\mu(f | \Lambda) + (f - E_\mu(f | \Lambda))$, and this decomposition inserted into the two integrals against g then shows that they are equal.

Thus, suppose conversely that

$$g := E_\lambda(f \circ \pi | \{\emptyset, X\} \otimes \Phi) \neq 0,$$

and hence

$$\int_Z (f \circ \pi) \cdot g \, d\lambda = \|g\|_2^2 \neq 0.$$

For each $i \in \mathbb{N}$ let $\alpha_i : Z' \rightarrow Y$ be the coordinate projection to the i^{th} copy of Y , and let $g_i := g \circ \alpha_i$. By the construction of λ' , the pair of coordinates $(\pi, \alpha_i) : Z' \rightarrow Z$ has the distribution λ for any i . This has the following two consequences:

- for any $M \geq 1$ one has

$$\int_{Z'} (f \circ \pi') \left(\frac{1}{M} \sum_{m=1}^M g_m \right) d\lambda' = \int_Z (f \circ \pi) \cdot g \, d\lambda = \|g\|_2^2 \neq 0;$$

- for all i one has

$$\mathbb{E}_{\lambda'}(g_i \mid \Sigma \otimes \{\emptyset, Y^{\mathbb{N}}\}) = \mathbb{E}_{\lambda'}(g_1 \mid \Sigma \otimes \{\emptyset, Y^{\mathbb{N}}\}),$$

so we may let h be this common conditional expectation.

Next, since all the Y -valued coordinates in Z' are relatively independent under λ' given the X -coordinate, one has

$$\int_{Z'} (g_i - h)(g_j - h) \, d\lambda' = 0 \quad \text{whenever } i \neq j,$$

and as $M \rightarrow \infty$ this implies the simple estimate

$$\left\| \frac{1}{M} \sum_{m=1}^M g_m - h \right\|_2^2 = \left\| \frac{1}{M} \sum_{m=1}^M (g_m - h) \right\|_2^2 = \frac{1}{M^2} \sum_{m=1}^M \|g_m - h\|_2^2 = O\left(\frac{1}{M}\right).$$

Hence

$$\frac{1}{M} \sum_{m=1}^M g_m \rightarrow h$$

in $\|\cdot\|_2$ as $M \rightarrow \infty$. On the one hand, this implies that h is a limit of functions measurable with respect to $\{\emptyset, X\} \otimes \Phi^{\otimes \mathbb{N}}$, hence is itself virtually measurable with respect to that σ -algebra. Therefore as a function on X it must actually be Λ -measurable. On the other hand, the above non-vanishing integral now gives

$$\int_{Z'} (f \circ \pi') \cdot h \, d\lambda' \neq 0.$$

Therefore $E_\mu(f | \Lambda) \neq 0$, so since Λ defines a C -factor of \mathbf{X} this completes the proof. \square

Remark This proof can be presented in several superficially different ways. On the one hand, it can be deduced almost immediately from a well-chosen appeal to the de Finetti-Hewitt-Savage Theorem, as in the paper [26] of Lesigne, Rittaud and de la Rue (see also Section 8.5 in Glasner [18]). On the other, it is a close cousin of the proof that for any idempotent class C , any system \mathbf{X} has an extension that is ‘ C -sated’ (Theorem 2.3.2 in [1]). \triangleleft

In previous applications, the idempotent classes of importance were those of the form $C_0^{H_1} \vee \dots \vee C_0^{H_\ell}$, introduced as examples above. Here we will need some slightly more complicated examples, because in order to account for the possible relations among the polynomials of a tuple \mathcal{F} we will need to consider simultaneously actions of G and also some ‘more free’ covering group $q : \tilde{G} \rightarrow G$.

Lemma 6.5 *Suppose that $q : H \rightarrow G$ is a continuous homomorphism of l.c.s.c. groups and that C is an idempotent class of H -systems. Then*

$$q_*C := \{G\text{-systems } \mathbf{X} \text{ such that } \mathbf{X}^{q(\cdot)} \in C\}$$

is an idempotent class of G -systems, and it is hereditary if C is hereditary.

Proof We must verify that q_*C is closed under joinings and inverse limits. Both are immediate: if \mathbf{Y} is a joining of $\mathbf{X}_i \in q_*C$ for $i = 1, 2$ then $\mathbf{Y}^{q(\cdot)}$ is the corresponding joining of $\mathbf{X}_i^{q(\cdot)}$, so lies in C because C is closed under joinings, and similarly for inverse limits. The last assertion also follows at once from the definition. \square

Definition 6.6 *The new class q_*C constructed in the previous lemma is the **image** of C under q .*

Lemma 6.7 *If C is an idempotent class of G -systems then*

$$\widehat{C} := \{\mathbf{X} : \mathbf{X} \text{ is a factor of a member of } C\}$$

is a hereditary idempotent class.

Proof The hereditary property is built into the definition, so once again it remains to check closure under joinings and inverse limits. Both are routine, so we give the

proof only for joinings. Suppose that $\mathbf{Y}_i = (Y_i, \Phi_i, \nu_i, v_i) \in \widehat{\mathbf{C}}$ for $i = 1, 2$, that $\pi_i : \mathbf{X}_i \rightarrow \mathbf{Y}_i$ are factors with $\mathbf{X}_i = (X_i, \Sigma_i, \mu_i, u_i) \in \mathbf{C}$ for $i = 1, 2$, and that $\mathbf{Z} = (Y_1 \times Y_2, \Phi_1 \otimes \Phi_2, \lambda, v_1 \times v_2)$ defines a joining of \mathbf{Y}_1 and \mathbf{Y}_2 . Then we may define a joining of \mathbf{X}_1 and \mathbf{X}_2 as a relatively independent product: letting $P_i : Y_i \rightarrow \text{Pr}(X_i)$ be a probability kernel representing the disintegration of μ_i over π_i , define

$$\lambda' := \int_{Y_1 \times Y_2} P(y_1, \cdot) \otimes P(y_2, \cdot) \lambda(dy_1, dy_2).$$

Now $(X_1 \times X_2, \Sigma_1 \otimes \Sigma_2, \lambda', u_1 \times u_2)$ is a joining of \mathbf{X}_1 and \mathbf{X}_2 , and hence a member of \mathbf{C} . The map $(x_1, x_2) \mapsto (\pi_1(x_1), \pi_2(x_2))$ witnesses \mathbf{Z} as a factor of this member of \mathbf{C} , so $\mathbf{Z} \in \widehat{\mathbf{C}}$. \square

Definition 6.8 *The class $\widehat{\mathbf{C}}$ constructed above is the **downward closure** of \mathbf{C} .*

When we come to apply this machinery, satedness relative only to classes of the form $\mathbf{C}_0^{H_1} \vee \dots \vee \mathbf{C}_0^{H_\ell}$ will not give us quite enough purchase over our situation. Instead we will need to first form an extended group $q : \tilde{G} \twoheadrightarrow G$ (in which copies of certain subgroups of G have been made ‘more independent’: see Section 8), and then for some subgroups $\tilde{H}_1, \tilde{H}_2, \dots, \tilde{H}_\ell \trianglelefteq \tilde{G}$ we will need to use satedness relative to the class

$$q_* \left((\mathbf{C}_0^{\tilde{H}_1} \vee \dots \vee \mathbf{C}_0^{\tilde{H}_\ell})^\wedge \right).$$

In prose, this is

‘The class of G -systems which, upon re-writing them as \tilde{G} -systems, become factors of joinings of systems in which one of the \tilde{H}_i acts trivially.’

This manoeuvre will appear during the proof of Proposition 8.2 below, where the need for it will become clearer. The particular way in which we will appeal to satedness with respect to such a class is captured by the following lemma.

Lemma 6.9 *Suppose that $q : H \twoheadrightarrow G$ is a continuous epimorphism of Lie groups, that \mathbf{C} is an idempotent class of H -systems and that $\mathbf{X} = (X, \Sigma, \mu, u)$ is a G -system. In addition, suppose that $f \in L^\infty(\mu)$ and that*

$$\pi : \mathbf{Y} = (Y, \Phi, \nu, v) \longrightarrow \mathbf{X}^{q(\cdot)}$$

is an extension of H -systems such that

$$E_\nu(f \circ \pi \mid C\Phi) \neq 0.$$

Then also

$$E_\mu(f \mid (q_*\widehat{C})\Sigma) \neq 0.$$

Proof We have $E_\nu(f \circ \pi \mid C\Phi) \neq 0$ by assumption, but on the other hand the function $f \circ \pi$ is invariant under v^h for every $h \in \ker q$:

$$f \circ \pi \circ v^h = f \circ u^{q(h)} \circ \pi = f \circ u^e \circ \pi = f \circ \pi.$$

Since $C\Phi$ is a factor of the whole H -action v , the conditional expectation operator $E_\nu(\cdot \mid C\Phi)$ preserves this $\ker q$ -invariance. Therefore $E_\nu(f \circ \pi \mid C\Phi)$ is measurable not only with respect to $C\Phi$ but also with respect to $\Phi^{\ker q}$.

Let $\alpha : \mathbf{Y} \longrightarrow \mathbf{Z}$ be a factor map onto another system which generates the factor $\Phi^{\ker q} \cap C\Phi \leq \Phi$, so its target system \mathbf{Z} is an element of \widehat{C} and has $\ker q$ acting trivially. Therefore this action of H may be identified with an action of G composed through q , say $\mathbf{Z} = \mathbf{W}^{q(\cdot)}$ for some G -system \mathbf{W} . (The joint measurability of v implies that of the action of G on \mathbf{W} , simply by choosing an everywhere-defined Borel selector $G \longrightarrow H$, as we clearly may for Lie group epimorphisms because they are locally diffeomorphic to orthogonal projections.)

Now the diagram

$$\begin{array}{ccc} & \mathbf{Y} & \\ \pi \swarrow & & \searrow \alpha \\ \mathbf{X}^{q(\cdot)} & & \mathbf{W}^{q(\cdot)} \end{array}$$

defines a joining of $\mathbf{X}^{q(\cdot)}$ and $\mathbf{W}^{q(\cdot)}$. It therefore also defines a joining of \mathbf{X} and \mathbf{W} , by simply identifying it with an invariant measure on $X \times W$ and writing the actions in terms of G rather than H .

Our assumption on f gives that $E(f \circ \pi \mid \alpha) \neq 0$. Therefore, within this joining of \mathbf{X} and \mathbf{W} , the lift of f has non-trivial conditional expectation onto the copy of \mathbf{W} , which is a member of $q_*\widehat{C}$, and so by Proposition 6.4 and Lemma 6.7 this implies $E_\mu(f \mid (q_*\widehat{C})\Sigma) \neq 0$. \square

7 The case of two-fold joinings

The case of Theorem 1.1 in which $k = 1$ will form the base of an inductive proof of the full theorem, and must be handled separately. Its proof is quite routine in the shadow of other works in this area, but it does already contain an appeal to the van der Corput estimate and an induction on the PET ordering for single polynomials (rather than whole tuples). It thus serves as a helpful preparation for the full induction that is to come.

Proposition 7.1 *Suppose that $\pi : G \curvearrowright \mathfrak{H}$ is an orthogonal representation and $\varphi : \mathbb{R} \times \mathbb{R}^r \rightarrow G$ is a polynomial map such that $\varphi(0, \cdot) \equiv e$. Then the operator averages*

$$\int_0^T \pi(\varphi(t, h)) dt$$

converge in the strong operator topology for every h , and the limit operator P_h is Zariski generically equal to the orthoprojection onto $\text{Fix}(\pi(\langle \text{img } \varphi \rangle))$.

Proof Step 1 First suppose that φ is linear in the first coordinate, meaning that $\varphi(\cdot, h)$ is a homomorphism for every $h \in \mathbb{R}^r$. Then for every h the map $t \mapsto \varphi(t, h)$ takes values in a 1-parameter subgroup of G , and so the classical ergodic theorem for orthogonal flows gives

$$\int_0^T \pi(\varphi(t, h)) dt \xrightarrow{\text{SOT}} P_h,$$

where P_h is the orthoprojection onto $\text{Fix}(\pi(\langle \text{img } \varphi(\cdot, h) \rangle))$. By Corollary 5.5 this equals $\text{Fix}(\pi(\langle \text{img } \varphi \rangle))$ Zariski generically, and so the proof is complete in the linear case.

Step 2 For arbitrary polynomial maps φ we show by PET induction that if

$$\int_0^T \pi(\varphi(t, h))v dt \not\rightarrow 0$$

for some $v \in \mathfrak{H}$, then $P_h v \neq 0$, where again P_h is the orthoprojection onto $\text{Fix}(\pi(\langle \text{img } \varphi(\cdot, h) \rangle))$. By decomposing an arbitrary v as $(1 - P_h)v + P_h v$ and appealing to Corollary 5.5 again, this will complete the proof.

If

$$\int_0^T \pi(\varphi(t, h))v dt \not\rightarrow 0$$

then the van der Corput estimate A.1 gives that also

$$\begin{aligned} \int_0^S \int_0^T \langle \pi(\varphi(t+s, h))v, \pi(\varphi(t, h))v \rangle dt ds \\ = \left\langle \int_0^S \int_0^T \pi(\varphi(t, h))^{-1} \varphi(t+s, h) v dt ds, v \right\rangle \not\rightarrow 0 \end{aligned}$$

as $T \rightarrow \infty$ and then $S \rightarrow \infty$.

By the special case of Lemma 4.13 for singleton families we have

$$\{(t, s, h) \mapsto \varphi(t, h)^{-1} \varphi(t+s, h)\} \prec_{\text{PET}} \{\varphi\},$$

and so the inductive hypothesis gives

$$\int_0^T \pi(\varphi(t, h))^{-1} \varphi(t+s, h) v dt \rightarrow Q_{s,h} v \quad \text{as } T \rightarrow \infty$$

with $Q_{s,h}$ the orthoprojection onto $\text{Fix}(\pi(\langle \text{img } \varphi(\cdot, h)^{-1} \varphi(\cdot+s, h) \rangle))$.

By Corollary 5.5, for every fixed h we have

$$\text{Fix}(\pi(\langle \text{img } \varphi(\cdot, h)^{-1} \varphi(\cdot+s, h) \rangle)) = \text{Fix}(\pi(\langle \text{img } \varphi(\cdot, h)^{-1} \varphi(\cdot+\cdot, h) \rangle))$$

for Zariski generic s , and now since $\varphi(0, h) \equiv e$ this is equal to

$$\text{Fix}(\pi(\langle \text{img } \varphi(\cdot, h) \rangle)).$$

In particular, for every h this equality must hold for Lebesgue-a.e. s , and thus our previous average over s may be written instead as

$$\int_0^S Q_{s,h} v ds = \int_0^S P_h v ds \equiv P_h v.$$

This proves that $P_h v \neq 0$, as required. \square

8 A partially characteristic factor

Now fix the following assumptions for this section and the next:

- G is an s -step connected and simply connected nilpotent Lie group;

- $\mathcal{F} = (\varphi_1, \varphi_2, \dots, \varphi_k)$ is a tuple of polynomial maps $\mathbb{R} \times \mathbb{R}^r \longrightarrow G$ with $k \geq 2$ in which φ_1 is a pivot, such that $\varphi_i(0, \cdot) \equiv e$ for each i , and such that $G = \langle \text{img } \varphi_1 \cup \dots \cup \text{img } \varphi_k \rangle$ (otherwise we may simply replace G with this smaller group);
- $(X_i, \Sigma_i, \mu_i, u_i)$ for $0 \leq i \leq k$ is a tuple of G -systems, and λ is a joining of them;
- A_T^λ for $T \in [0, \infty)$ is the family of averaging operators associated to the orbit of λ under $(\varphi_1, \varphi_2, \dots, \varphi_k)$ as in Theorem 1.3, so note that these implicitly depend on h , the parameter in the argument of the φ_i which is *not* averaged.

At the heart of the inductive proof of Theorem 1.3 lies a result promising that in order to study the functional averages $A_T^\lambda(f_1, f_2, \dots, f_k)$, one may assume that one of the functions f_i has some special additional structure (which we will see later enables a further reduction to the case of a simpler family of polynomial maps). This extra structure is captured by a simple adaptation of an important idea introduced in [17], and which has been used extensively since (see, for instance, [21, 44, 5, 3]).

Definition 8.1 (Partially characteristic factor) *In the above setting a factor $\Lambda \leq \Sigma_1$ is **partially characteristic** for the averages A_T^λ if for any tuple of functions $f_i \in L^\infty(\mu_i)$ one has*

$$\|A_T^\lambda(f_1, f_2, \dots, f_k) - A_T^\lambda(\mathbb{E}(f_1 | \Lambda), f_2, \dots, f_k)\|_2 \longrightarrow 0$$

as $T \longrightarrow \infty$ for Zariski generic h (recalling that the operators A_T^λ implicitly depend on $h \in \mathbb{R}^r$).

Remark The main difference between this definition and its predecessors in earlier papers is that here, in consonance with the statement of Theorem 1.3, we require convergence only for Zariski generic h .

As stated, this definition allows the Zariski meagre set $F \subseteq \mathbb{R}^r$ containing those h for which convergence fails to depend on f_1, f_2, \dots, f_k . However, it is easily checked that for a given h , this convergence holds for all tuples of functions if one knows that it holds for tuples drawn from some $\|\cdot\|_2$ -dense subsets of the unit balls of $L^\infty(\mu_i)$, $i = 1, 2, \dots, k$. Since one can choose countable such subsets, we deduce that there is a countable intersection of Zariski residual subsets of \mathbb{R}^r (which is therefore still Zariski residual) on which the above convergence holds for all tuples of functions. \triangleleft

As in many of the earlier works cited above, the first step towards proving the convergence of $A_T^\lambda(f_1, \dots, f_k)$ will be to identify a partially characteristic factor with some useful structure. However, a new twist appears in the present setting: here we must first pass from G -systems to actions of some covering group of G .

To be precise, let

$$\tilde{\varphi}_1 : (t, h) \mapsto (\varphi_1(t, h), \dots, \varphi_k(t, h)),$$

let

$$\tilde{\varphi}_i : (t, h) \mapsto (\varphi_i(t, h), \dots, \varphi_i(t, h)), \quad \text{for } i = 2, 3, \dots, k,$$

(notice the subscripts in different coordinates), and let

$$\tilde{G} := \langle \text{img } \tilde{\varphi}_1 \cup \text{img } \tilde{\varphi}_2 \cup \dots \cup \text{img } \tilde{\varphi}_k \rangle \leq G^{k+1}.$$

Let $q : \tilde{G} \rightarrow G$ be the restriction to \tilde{G} of the projection $G^k \rightarrow G$ onto the first coordinate. Then q intertwines each $\tilde{\varphi}_i$ with φ_i for $i \geq 1$ (because φ_i appears in the first coordinate of $\tilde{\varphi}_i$ for every i).

It is easy to verify that $q(\tilde{G}) = G$. The group \tilde{G} is connected, because each $\tilde{\varphi}_i(\cdot, h)$ passes through the origin for every h , and hence $\tilde{G} = \exp V$ for some Lie subalgebra $V \leq \mathfrak{g}^k$. The image of V under the first coordinate projection is a Lie subalgebra $V_1 \leq \mathfrak{g}$, and since G is simply connected it follows that $\exp V_1$ is a closed subgroup of G which is contained in $q(\tilde{G})$. On the other hand it must contain $\text{img } \varphi_i$ for every $i \leq k$, so in fact $q(\tilde{G}) = \exp V_1 = G$.

The next technical proposition lies at the heart of all that follows. It provides a partially characteristic factor of $\mathbf{X}_1 = (X_1, \Sigma_1, \mu_1, u_1)$ for the averages A_T^λ , but only at the cost of regarding instead the modified system $\mathbf{X}_1^{q(\cdot)}$. The need for this sleight of hand will become clear during the proof.

Proposition 8.2 *Assume that conclusions (1–3) of Theorem 1.3 have already been established for all polynomial families preceding \mathcal{F} in the PET ordering, suppose that φ_1 is a pivot, and let*

$$\mathbf{C} := q_* \left(\left(\mathbf{C}_0^{\langle \text{img } \tilde{\varphi}_1 \rangle} \vee \bigvee_{j=2}^k \mathbf{C}_0^{\langle \text{img } \tilde{\varphi}_j \tilde{\varphi}_1^{-1} \rangle} \right)^\wedge \right).$$

(Recall the discussion following Definition 6.8.) Then for any systems \mathbf{X}_i , $i = 0, 1, \dots, k$, the factor $\mathbf{C}\Sigma_1 \leq \Sigma_1$ is partially characteristic.

Remark Of course, once this proposition has been proved then it implies some conclusion even if $A_T^\lambda(f_1, \dots, f_k) \not\rightarrow 0$ for just one value of h , because by fixing that h we may simply regard each φ_i as a polynomial function of t alone, and so apply the proposition with $r = 0$. Indeed, we will use this trick a few times later. However, one must beware of the delicacy that the idempotent class appearing in this proposition may not be the same after one makes such a restriction, so nor will the σ -sigma algebra $C\Sigma_1$ in general. Even the group extension $q : \tilde{G} \rightarrow G$ itself will not be the same as above, but will depend on the choice of h . Since at some points later we will really need the above conclusion about the generic behaviour of the averages in h , it seems easiest to formulate it as here and then apply it with a restricted parameter space when convenient. \triangleleft

Proof Since any f_1 may be decomposed as

$$E_\mu(f_1 | C\Sigma) + (f_1 - E_\mu(f_1 | C\Sigma))$$

and the operator A_T^λ is multilinear, it is enough to prove that if $E_\mu(f_1 | C\Sigma) = 0$ then for any f_2, \dots, f_k one has

$$\|A_T^\lambda(f_1, f_2, \dots, f_k)\|_2 \rightarrow 0$$

as $T \rightarrow \infty$ for Zariski generic h . Contrapositively, this is equivalent to showing that if the set

$$E := \{h \in \mathbb{R}^r : \|A_T^\lambda(f_1, f_2, \dots, f_k)\|_2 \not\rightarrow 0 \text{ as } T \rightarrow \infty\}$$

is not Zariski meagre then $E_\mu(f_1 | C\Sigma) \neq 0$. Henceforth we assume that E is not Zariski meagre.

Furthermore, in view of Lemma 6.9, it now suffices to find an extension of spaces $\pi : (\tilde{X}, \tilde{\Sigma}, \tilde{\mu}) \rightarrow (X_1, \Sigma_1, \mu_1)$ and an action $\tilde{u} : \tilde{G} \curvearrowright (\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$ such that $\pi \circ \tilde{u} = u^{q(\cdot)}$ and

$$E(f_1 \circ \pi | \Lambda) \neq 0,$$

where

$$\Lambda := \tilde{\Sigma} \langle \text{img } \tilde{\varphi}_1 \rangle \vee \bigvee_{i=2}^k \tilde{\Sigma} \langle \text{img } \tilde{\varphi}_1 \cdot \tilde{\varphi}_i^{-1} \rangle.$$

This is the point at which we have made use of the general properties of idempotent classes. This implication will follow in two steps: applying the van der Corput estimate (Lemma A.1), and interpreting what it tells us.

Step 1 Letting

$$g_{t,h} := M^\lambda(f_1 \circ u_1^{\varphi_1(t,h)}, f_2 \circ u_2^{\varphi_2(t,h)}, \dots, f_k \circ u_k^{\varphi_k(t,h)}),$$

the van der Corput estimate implies that for $h \in E$ one also has

$$\int_0^S \int_0^T \int_{X_0} g_{t+s,h} g_{t,h} d\mu_0 dt ds \not\rightarrow 0$$

as $T \rightarrow \infty$ and then $S \rightarrow \infty$.

For each s , by Lemma 2.1 we may re-write the two inner integrals here as

$$\begin{aligned} \int_0^T \int_{X_1^2 \times \dots \times X_k^2} & (f_1 \circ u_1^{\varphi_1(t,h)}) \otimes (f_1 \circ u_1^{\varphi_1(s,h)\psi_1(t,s,h)}) \otimes \\ & \dots \otimes (f_k \circ u_k^{\varphi_k(t,h)}) \otimes (f_k \circ u_k^{\varphi_k(s,h)\psi_k(t,s,h)}) d(\lambda \otimes_0 \lambda) dt, \end{aligned}$$

where

$$\psi_i(t, s, h) := \varphi_i(s, h)^{-1} \varphi_i(t + s, h) \quad \text{for each } i = 1, 2, \dots, k,$$

so $\psi_i : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^r \rightarrow G$ is a polynomial map with the property that $\psi_i(0, \cdot, \cdot) \equiv e$.

Since $\lambda \otimes_0 \lambda$ is a joining of two duplicates of each of the G -systems $(X_i, \Sigma_i, \mu_i, u_i)$ for $1 \leq i \leq k$, it is invariant under the diagonal transformations $u_{\Delta}^{\varphi_1(t,h)^{-1}}$. Applying this within the above integral shows that it is equal to

$$\begin{aligned} \int_0^T \int_{X_1^2 \times \dots \times X_k^2} & f_1 \otimes (f_1 \circ u_1^{\varphi_1(s,h)\psi'_1(t,s,h)}) \otimes \\ & \dots \otimes (f_k \circ u_k^{\varphi'_k(t,h)}) \otimes (f_k \circ u_k^{\varphi_k(s,h)\psi'_k(t,s,h)}) d(\lambda \otimes_0 \lambda) dt \end{aligned}$$

with

$$\begin{aligned} \psi'_i(t, s, h) &:= \psi_i(t, s, h) \varphi_1(t, h)^{-1} \quad \text{for } i \geq 1 \text{ and} \\ \varphi'_i(t, h) &:= \varphi_i(t, h) \varphi_1(t, h)^{-1} \quad \text{for } i \geq 2. \end{aligned}$$

We recognize these as comprising the 1st derived family of \mathcal{F} , which by Lemma 4.13 precedes \mathcal{F} in the PET ordering because φ_1 was a pivot. Let

$$\vec{\psi}: (t, s, h) \mapsto (e, \psi'_1(t, s, h), \varphi'_2(t, h), \psi'_2(t, s, h), \dots, \varphi'_k(t, h), \psi'_k(t, s, h)).$$

By the inductive hypothesis, for every $h \in \mathbb{R}^r$ there are a Zariski residual set $F_h \subseteq \mathbb{R}$ and a joining θ^h on $X_1^2 \times X_2^2 \times \dots \times X_k^2$ invariant under

$$\langle G^{\Delta 2k} \cup \text{img } \vec{\psi}(\cdot, \cdot, h) \rangle$$

such that for all $s \in F_h$ the above integral tends to

$$\int_{X_1^2 \times \dots \times X_k^2} f_1 \otimes (f_1 \circ u_1^{\varphi_1(s,h)}) \otimes \dots \otimes f_k \otimes (f_k \circ u_k^{\varphi_k(s,h)}) d\theta^h$$

as $T \rightarrow \infty$. Moreover these θ^h are equal to one fixed joining θ on a Zariski residual set of h , so that this θ must in fact be invariant under $\langle G^{\Delta_{2k}} \cup \text{img } \overrightarrow{\psi} \rangle$.

Since the Zariski residual set F_h has full Lebesgue measure, for each h our previous average over s may now be replaced by

$$\int_0^S \int_{X_1^2 \times \dots \times X_k^2} f_1 \otimes (f_1 \circ u_1^{\varphi_1(s,h)}) \otimes \dots \otimes f_k \otimes (f_k \circ u_k^{\varphi_k(s,h)}) d\theta^h ds,$$

implying that for $h \in E$ this also does not vanish as $S \rightarrow \infty$.

Next, one has

$$\begin{aligned} & (\varphi_1(s, h), \varphi_1(s, h), \dots, \varphi_k(s, h), \varphi_k(s, h)) \\ &= (e, e, \dots, \varphi_k(s, h)\varphi_1(s, h)^{-1}, \varphi_k(s, h)\varphi_1(s, h)^{-1}) \\ & \quad \cdot (\varphi_1(s, h), \varphi_1(s, h), \dots, \varphi_1(s, h), \varphi_1(s, h)) \\ &= \overrightarrow{\psi}(s, 0, h) \cdot (\varphi_1(s, h), \varphi_1(s, h), \dots, \varphi_1(s, h), \varphi_1(s, h)) \\ &\in \langle G^{\Delta_{2k}} \cup \text{img } \overrightarrow{\psi}(\cdot, \cdot, h) \rangle \end{aligned}$$

for every s , and so each joining θ^h is already invariant under the new off-diagonal polynomial flow

$$\xi(\cdot, h) : s \mapsto (\varphi_1(s, h), \varphi_1(s, h), \dots, \varphi_k(s, h), \varphi_k(s, h)).$$

Since we may re-write the above average as

$$\int_0^S \int_{X_1^2 \times \dots \times X_k^2} (f_1 \otimes 1 \otimes \dots \otimes f_k \otimes 1) \cdot ((1 \otimes f_1 \otimes \dots \otimes 1 \otimes f_k) \circ u_{\times}^{\xi(s,h)}) d\theta^h ds,$$

by the base case Proposition 7.1 it must converge to

$$\int_{X_1^2 \times \dots \times X_k^2} (f_1 \otimes 1 \otimes \dots \otimes f_k \otimes 1) \cdot \mathbb{E}(1 \otimes f_1 \otimes \dots \otimes 1 \otimes f_k \mid \Sigma_{\times}^{\langle \text{img } \xi(\cdot, h) \rangle}) d\theta^h$$

as $S \rightarrow \infty$, where $\Sigma_{\times} := \Sigma_1^{\otimes 2} \otimes \dots \otimes \Sigma_k^{\otimes 2}$ and the conditional expectation here is with respect to θ^h .

Therefore this last integral is nonzero for every $h \in E$. Since the sets

$$\{h : \theta^h \neq \theta\} \quad \text{and} \quad \{h : \Sigma_{\times}^{\langle \text{img } \xi(\cdot, h) \rangle} \neq \Sigma_{\times}^{\langle \text{img } \xi \rangle} \text{ up to } \theta\text{-negligible sets}\}$$

both are Zariski meagre (the latter by Corollary 5.5), their union cannot contain E , and so any value $h \in E$ that is not in either of these meagre sets witnesses that

$$\int_{X_1^2 \times \dots \times X_k^2} (f_1 \otimes 1 \otimes \dots \otimes f_k \otimes 1) \cdot E(1 \otimes f_1 \otimes \dots \otimes 1 \otimes f_k \mid \Sigma_{\times}^{\langle \text{img } \xi \rangle}) d\theta \neq 0.$$

Step 2 Now set

$$(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}) := \left(\prod_{i=1}^k X_i^2, \bigotimes_{i=1}^k \Sigma_i^{\otimes 2}, \theta \right)$$

and let $\pi : \tilde{X} \rightarrow X_1$ be the coordinate projection onto the first copy of X_1 . Observe that the polynomial map ξ defined in Step 1 is simply a copy of $\tilde{\varphi}_1$ in which each coordinate has been duplicated. Define $q_1 : \tilde{G} \rightarrow G^{2k}$ to be the restriction to \tilde{G} of the coordinate-duplicating map

$$(g_1, g_2, \dots, g_k) \mapsto (g_1, g_1, g_2, g_2, \dots, g_k, g_k).$$

Composing q_1 with the Cartesian product action u_{\times} of G^{2k} now gives an action \tilde{u} of \tilde{G} on $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu})$, since we have already deduced from our inductive hypotheses that $\tilde{\mu} = \theta$ is invariant under u_{Δ} (and hence the image of $q_1 \circ \tilde{\varphi}_i$ for each $i \geq 2$) and also under $\langle \text{img } \xi \rangle$ (which is the image of $q_1 \circ \tilde{\varphi}_1$).

On the first coordinate in $\prod_{i=1}^k X_i^2$, the transformation \tilde{u}^g simply agrees with u^g for any $g \in \langle \text{img } \tilde{\varphi}_2 \cup \dots \cup \text{img } \tilde{\varphi}_k \rangle$. On the other hand,

$$\pi \circ \tilde{u}^{\tilde{\varphi}_1(t, h)} \stackrel{\text{def}}{=} \pi \circ (u_1^{\varphi_1(t, h)} \times u_1^{\varphi_1(t, h)} \times \dots \times u_k^{\varphi_k(t, h)} \times u_k^{\varphi_k(t, h)}) = u_1^{\varphi_1(t, h)}.$$

Since these cases together generate the whole of \tilde{G} , it follows that $\pi \circ \tilde{u}^{\tilde{g}} = u^{q(\tilde{g})}$ for all $\tilde{g} \in \tilde{G}$, where $q : \tilde{G} \rightarrow G$ is the covering homomorphism constructed previously.

Finally, an inspection of the action \tilde{u} on the other coordinates of \tilde{X} shows that

- for each $i \in \{2, 3, \dots, k\}$ the transformations $\tilde{u}^{\tilde{\varphi}_1(t, h)}$ and $\tilde{u}^{\tilde{\varphi}_i(t, h)}$ agree on the first coordinate copy of X_i , and
- the function $E(1 \otimes f_1 \otimes \dots \otimes 1 \otimes f_k \mid \Sigma_{\times}^{\langle \text{img } \xi \rangle})$ is invariant under the \tilde{u} -action of $\langle \text{img } \tilde{\varphi}_1 \rangle$.

Therefore the non-vanishing of the integral at the end of step 1 asserts that $f_1 \circ \pi$ has a non-zero inner product with a function that is manifestly measurable with respect to a system in the class

$$C_0^{\langle \text{img } \tilde{\varphi}_1 \rangle} \vee \bigvee_{j=2}^k C_0^{\langle \text{img } \tilde{\varphi}_j \tilde{\varphi}_1^{-1} \rangle},$$

and hence $E(f_1 \circ \pi \mid C\Sigma_1) \neq 0$, as required. \square

Remarks 1. The above proof makes clear the need to extend the modified system $\mathbf{X}_1^{q(\cdot)}$, rather than \mathbf{X}_1 itself. We constructed our extension from some joining on $X_1^2 \times \cdots \times X_k^2$ through the coordinate projection onto X_1 , and in order to derive the desired nonzero conditional expectation for it we needed the polynomial trajectory of transformations $u_1^{\varphi_1(t,h)}$ downstairs to lift to the trajectory

$$u_1^{\varphi_1(t,h)} \times u_1^{\varphi_1(t,h)} \times u_2^{\varphi_2(t,h)} \times u_2^{\varphi_2(t,h)} \times \cdots \times u_k^{\varphi_k(t,h)} \times u_k^{\varphi_k(t,h)}.$$

The new map $\tilde{\varphi}_1$ may not be a PET-minimal member of $(\tilde{\varphi}_1, \dots, \tilde{\varphi}_k)$, and it also may not share its leading term with any of the lifts $\tilde{\varphi}_i$ for $i \geq 2$, even if φ_1 downstairs does have some leading terms in common with the other φ_i . Thus in order to write these $\tilde{\varphi}_i$ as genuine lifts of the φ_i we must first split the group G apart slightly in order to separate these leading terms. Happily, the problem itself gives us a natural way to do this: the lifted polynomial mapping $\tilde{\varphi}_1$ is suitably ‘separated’ from $\tilde{\varphi}_i$, $i \geq 2$, inside the Cartesian product G^k , so we have simply taken \tilde{G} to be the closed subgroup of G^k generated by these lifted mappings and composed our actions with the quotient map $q : \tilde{G} \rightarrow G$.

2. If a factor $\Lambda \leq \Sigma_1$ is partially characteristic and we assume that the limits $\lambda^h = \lim_{T \rightarrow \infty} \lambda_T^h$ exist, then considering the integral formula

$$\int_{\prod_i X_i} f_0 \otimes f_1 \otimes \cdots \otimes f_k \, d\lambda_T^h = \int_{X_0} f_0 \cdot A_T^\lambda(f_1, \dots, f_k) \, d\mu_0$$

shows that for Zariski generic h the coordinate projection $\prod_i X_i \rightarrow X_1$ is relatively independent under λ^h over its further factor generated by $\Lambda \leq \Sigma_1$. Thus, knowledge of a non-trivial partially characteristic factor gives some structural information about the limit joining.

In particular, consider a case in which q is an isomorphism (so that the subgroups $\langle \text{img } \varphi_1 \rangle$ and $\langle \text{img } \varphi_2 \cup \cdots \cup \text{img } \varphi_k \rangle$ are already sufficiently ‘spread apart’ in G), and suppose furthermore that the factor $C\mathbf{X}_1$ can itself be expressed as a joining of systems $\mathbf{Z}_0 \in C_0^{\langle \text{img } \varphi_1 \rangle}$ and $\mathbf{Z}_i \in C_0^{\langle \text{img } (\varphi_1 \varphi_i^{-1}) \rangle}$ for $i \geq 2$ (rather than just

as a factor of such). Then we know that any limit joining λ' must be relatively independent over the factor CX_1 , and upon restricting ourselves to this factor we can express λ' alternatively as a joining of

$$X_0, Z_0, Z_2, \dots, Z_k, X_2, \dots, X_k.$$

(In fact we will use a similar manipulation in the next section). Moreover, the assumption that CX_1 itself be a joining is not terribly restrictive, since an arbitrary system X_1 always has an extension for which this is true (by using the machinery of ‘C-sated’ extensions, as developed in Chapter 2 of [1]).

It would be interesting to know whether further use of the ideas behind Proposition 8.2 could give a more complete picture of the possible structure of λ' . This would presumably involve repeated assertions of relative independence over increasingly ‘small’ factors of the original system, on which increasingly large subgroups of G act trivially. Such a picture does emerge in the study of the linear multiple averages constructed from a tuple of \mathbb{Z}^d -actions (see Chapter 4 of [1]), but in the present setting the need to keep track of a large family of different subgroups of G may make the resulting description more obscure.

Even without a manageable description, this kind of result suggests that the limit joining λ' of Theorem 1.1 not only exists, but exhibits some rigidity over different possible initial joinings λ , since λ' must exhibit these various instances of relative independence. Once again there is a superficial analogy here with the study of unipotent flows on homogeneous spaces, where a central theme is the classification of all possible invariant measures and the rigidity that such a classification entails; but once again, I do not know whether this points to any deeper connexions between that setting and ours. \triangleleft

9 Proof of the main theorem

We can now complete the proof of Theorem 1.3. The general case is handled by a ‘spiral’ PET induction on the tuple $(\varphi_1, \dots, \varphi_k)$: for each such tuple we will show that

$$\begin{aligned} & (\text{assertions (1,2,3) for } (\tilde{\varphi}_2, \dots, \tilde{\varphi}_k)) \\ & \Rightarrow (\text{assertion (1) for } (\varphi_1, \varphi_2, \dots, \varphi_k)) \\ & \Rightarrow (\text{assertion (2) for } (\varphi_1, \varphi_2, \dots, \varphi_k)) \\ & \Rightarrow (\text{assertion (3) for } (\varphi_1, \varphi_2, \dots, \varphi_k)), \end{aligned}$$

at which point the induction closes on itself.

We retain the assumptions from the start of Section 8. Proposition 8.2 gives the purchase needed to complete our induction. Let the class \mathbf{C} and group extension $q : \tilde{G} \longrightarrow G$ be as in the preceding section. In analysing the family of averages

$$A_T^\lambda(f_1, f_2, \dots, f_k),$$

Proposition 8.2 allows us to assume that f_1 is measurable with respect to the factor $\mathbf{C}\Sigma_1$, or equivalently that \mathbf{X}_1 is itself a system with the property that the \tilde{G} -system $\mathbf{X}_1^{q(\cdot)}$ is a factor of a member of the class

$$\mathbf{C}_0^{\langle \text{img } \tilde{\varphi}_1 \rangle} \vee \bigvee_{j=2}^k \mathbf{C}_0^{\langle \text{img } \tilde{\varphi}_j \tilde{\varphi}_1^{-1} \rangle}.$$

From this point a careful re-arrangement gives a reduction to the conclusions of Theorem 1.2 for the group \tilde{G} and family $(\tilde{\varphi}_2, \dots, \tilde{\varphi}_k)$, which is isomorphic to $(\varphi_2, \dots, \varphi_k)$ and hence precedes $(\varphi_1, \varphi_2, \dots, \varphi_k)$ in the PET ordering (see Lemma 4.13). Note that this holds in spite of our ascent from G to \tilde{G} , because we have now removed $\tilde{\varphi}_1$ from the picture altogether.

In order to set up the necessary re-arrangement, assume that $\mathbf{X}_1 = \mathbf{C}\mathbf{X}_1$. By the definition of \mathbf{C} there are a system $\tilde{\mathbf{X}} \in \mathbf{C}_0^{\langle \text{img } \tilde{\varphi}_1 \rangle} \vee \bigvee_{j=2}^k \mathbf{C}_0^{\langle \text{img } \tilde{\varphi}_j \tilde{\varphi}_1^{-1} \rangle}$ and a factor map $\pi : \tilde{\mathbf{X}} \longrightarrow \mathbf{X}_1^{q(\cdot)}$.

Now let $\tilde{\mathbf{X}}_1 := \tilde{\mathbf{X}}$ and $\tilde{\mathbf{X}}_i := \mathbf{X}_i^{q(\cdot)}$ for any $i \neq 1$, and choose any lift of λ to a joining $\tilde{\lambda}$ of the $\tilde{\mathbf{X}}_i$ (for instance, one could use the relatively independent product over λ). For each $i \neq 1$ consider the factor $\tilde{\Sigma}_1^{\langle \text{img } \tilde{\varphi}_1 \tilde{\varphi}_i^{-1} \rangle} \leq \tilde{\Sigma}_1$, and let

$$\zeta_i : \tilde{X}_1 \longrightarrow Z_i$$

be a factor map of standard Borel \tilde{G} -space which generates this factor. These may be realized as factors of the joining $\tilde{\lambda}$ through the coordinate projection $\prod_i \tilde{X}_i \longrightarrow \tilde{X}_1$. Crucially, by enlarging each of the systems $\tilde{\mathbf{X}}_i$ for $i \neq 1$, we can arrange that under $\tilde{\lambda}$ each of these factor maps to Z_i is also virtually measurable with respect to the \tilde{X}_i -coordinate, as well as the \tilde{X}_1 -coordinate. To this end, for each $i \neq 1$ consider the composition

$$\tilde{X}_0 \times \dots \times \tilde{X}_k \xrightarrow{\text{coord. proj.}} \tilde{X}_1 \times \tilde{X}_i \xrightarrow{\zeta_i \times \text{id}} Z_i \times \tilde{X}_i.$$

Since this composition respects the \tilde{G} -actions, it defines a joining of \mathbf{Z}_i with $\tilde{\mathbf{X}}_i$, which we denote by $\mathbf{Y}_i = (Y_i, \Phi_i, \nu_i, v_i)$. Let $\eta_i : Y_i \longrightarrow \tilde{X}_i$ be the second coordinate projection.

Thus we have constructed a collection of factorizations

$$\begin{array}{ccc}
 (\tilde{X}_0 \times \cdots \times \tilde{X}_k, \tilde{\Sigma}_0 \otimes \cdots \otimes \tilde{\Sigma}_k, \tilde{\lambda}, \tilde{u}_\Delta) & \xrightarrow{\text{coord. proj.}} & (\tilde{X}_i, \tilde{\Sigma}_i, \tilde{\mu}_i, \tilde{u}_i) \\
 & \searrow \quad \swarrow & \\
 & (Y_i, \Phi_i, \nu_i, v_i) &
 \end{array}$$

for each $i \in \{0, 2, 3, \dots, k\}$. Putting these together with the coordinate projection $\tilde{X}_0 \times \cdots \times \tilde{X}_k \longrightarrow \tilde{X}_1$ therefore gives a measure-theoretic isomorphism

$$\begin{aligned}
 & (\tilde{X}_0 \times \cdots \times \tilde{X}_k, \tilde{\Sigma}_0 \otimes \cdots \otimes \tilde{\Sigma}_k, \tilde{\lambda}, \tilde{u}_\Delta) \\
 & \xrightarrow{\cong} (Y_0 \times \tilde{X}_1 \times Y_2 \times \cdots \times Y_k, \Phi_0 \otimes \tilde{\Sigma}_1 \otimes \Phi_2 \otimes \cdots \otimes \Phi_k, \theta, v_\Delta)
 \end{aligned}$$

for some joining θ of \tilde{G} -systems.

In addition, this construction guarantees that the factor maps

$$\tilde{X}_0 \times \cdots \times \tilde{X}_k \xrightarrow{\text{coord. proj.}} \tilde{X}_1 \xrightarrow{\zeta_i} Z_i$$

and

$$\tilde{X}_0 \times \cdots \times \tilde{X}_k \longrightarrow Y_i \xrightarrow{\text{coord. proj.}} Z_i$$

agree up to λ' -negligible sets. Therefore any $h \in L^\infty(\tilde{\mu}_1)$ which is measurable with respect to $\tilde{\Sigma}_1^{\langle \text{img } \tilde{\varphi}_1 \tilde{\varphi}_i^{-1} \rangle}$ (equivalently, which is invariant under $\langle \text{img } \tilde{\varphi}_1 \tilde{\varphi}_i^{-1} \rangle$, with the convention that $\tilde{\varphi}_0 \equiv e$) has an essentially unique counterpart $h' \in L^\infty(\nu_i)$ which lifts to the same function on $\tilde{X}_0 \times \cdots \times \tilde{X}_k$ up to $\tilde{\lambda}$ -negligible sets, and which is invariant under the same subgroup of \tilde{G} .

Lemma 9.1 *In the situation described above, consider the averaging operators associated to the lifted family of polynomial maps $\tilde{\varphi}_i : \mathbb{R} \times \mathbb{R}^r \longrightarrow \tilde{G}$. Suppose that $f_1 \in L^\infty(\tilde{\mu}_1)$ is a function of the special form*

$$g \cdot h_2 \cdot \cdots \cdot h_k,$$

where $g \in L^\infty(\tilde{\mu}_1)$ is invariant under $\langle \text{img } \tilde{\varphi}_1 \rangle$ and each h_i is invariant under $\langle \text{img } \tilde{\varphi}_1 \tilde{\varphi}_i^{-1} \rangle$. Then for any other functions $f_i \in L^\infty(\tilde{\mu}_i)$ for $i \neq 1$ one has

$$A_T^{\tilde{\lambda}}(f_1, f_2, \dots, f_k) = \mathbb{E}(g' \cdot A_T^\theta(1, h'_2(f_2 \circ \eta_2), \dots, h'_k(f_k \circ \eta_k)) \mid \eta_0)$$

(recalling that $A_T^{\tilde{\lambda}}$ has range in $L^\infty(\tilde{\mu}_0)$, while A_T^θ has range in $L^\infty(\tilde{\nu}_0)$), where g' and h'_i are the counterparts of g and h_i introduced above.

Proof By the definition of $A_T^{\tilde{\lambda}}$ and A_T^θ this follows from the analogous calculation at the level of joinings. For the joinings $\tilde{\lambda}$ and θ , the above isomorphism gives

$$\begin{aligned} & \int_0^T \int_{\prod_i \tilde{X}_i} f_0 \otimes (f_1 \circ \tilde{u}_1^{\tilde{\varphi}_1(t,h)}) \otimes \cdots \otimes (f_k \circ \tilde{u}_k^{\tilde{\varphi}_k(t,h)}) d\tilde{\lambda} dt \\ &= \int_0^T \int_{Y_0 \times \tilde{X}_1 \times Y_2 \times \cdots \times Y_k} (f_0 \circ \eta_0) \otimes (f_1 \circ \tilde{u}_1^{\tilde{\varphi}_1(t,h)}) \\ & \quad \otimes (f_2 \circ \eta_2 \circ v_2^{\tilde{\varphi}_2(t,h)}) \otimes \cdots \otimes (f_k \circ \eta_k \circ v_k^{\tilde{\varphi}_k(t,h)}) d\theta dt. \end{aligned}$$

On the other hand, our assumptions on the structure of f_1 imply that

$$g \circ \tilde{u}_1^{\tilde{\varphi}_1(t,h)} = g \quad \text{and} \quad h_i \circ \tilde{u}_1^{\tilde{\varphi}_1(t,h)} = h_i \circ \tilde{u}_1^{\tilde{\varphi}_i(t,h)}, \quad i = 2, 3, \dots, k,$$

for all (t, h) . Also, the counterparts $g' \in L^\infty(\nu_0)$ and $h'_i \in L^\infty(\nu_i)$ for $i \geq 2$ satisfy

$$g'(y_0) = g(\tilde{x}_1) \quad \text{and} \quad h'_i(y_i) = h_i(\tilde{x}_1)$$

for θ -almost every $(y_0, \tilde{x}_1, y_2, \dots, y_k)$. The above integral with respect to θ may therefore be re-written as

$$\begin{aligned} & \int_0^T \int_{Y_0 \times \tilde{X}_1 \times Y_2 \times \cdots \times Y_k} (g'(f_0 \circ \eta_0)) \otimes 1_{\tilde{X}_1} \otimes ((h'_2(f_2 \circ \eta_2)) \circ v_2^{\tilde{\varphi}_2(t,h)}) \otimes \\ & \quad \cdots \otimes ((h'_k(f_k \circ \eta_k)) \circ v_k^{\tilde{\varphi}_k(t,h)}) d\theta dt. \end{aligned}$$

Regarded as a linear functional applied to f_0 , this is integration against

$$E(g \cdot A_T^\theta(1, h_2(f_2 \circ \eta_2), \dots, h_k(f_k \circ \eta_k)) \mid \eta_0),$$

as required. □

Of course, the importance of the above lemma is that on the right-hand side there is no non-trivial function in the first entry under A_T^θ . This now leads quite smoothly to a completion of our spiral induction.

Proof of Theorem 1.3 In case $k = 1$, M^λ extends to a bounded operator $L^2(\mu_1) \rightarrow L^2(\mu_0)$ and the desired assertions of convergence and genericity become simply that (i) the average

$$M^\lambda \left(\int_0^T u(\varphi_1(t, h)^{-1})^* f_1 dt \right)$$

converges to $M^\lambda P_h f$ with P_h the conditional expectation onto $\Sigma_1^{\langle \text{img } \varphi(\cdot, h) \rangle}$, and (ii) this is generically equal to $M^\lambda P f$ with P the conditional expectation onto $\Sigma_1^{\langle \text{img } \varphi \rangle}$. Both of these assertions follow at once from Proposition 7.1.

It remains to handle the inductive step in case $k \geq 2$. Assume that properties (1–3) have already been proved for all tuples preceding \mathcal{F} in the PET ordering. We will deduce those properties for \mathcal{F} in order.

Property (1) In this step, by fixing one h throughout the proof and replacing G with its subgroup $\langle \text{img } \varphi_1(\cdot, h) \cup \dots \cup \text{img } \varphi_k(\cdot, h) \rangle$ if necessary, we may assume that each φ_i is a function of t alone, and hence that $r = 0$. With this agreed, let $q : \tilde{G} \rightarrow G$ and the class C be constructed as before using this new group and tuple of maps.

By re-ordering \mathcal{F} if necessary we may also assume φ_1 is a pivot. In this case, by Proposition 8.2 it suffices to show that the averages $A_T^\lambda(f_1, \dots, f_k)$ converge when f_1 is $C\Sigma_1$ -measurable.

Construct the \tilde{G} -systems $\tilde{\mathbf{X}}_i$ and \mathbf{Y}_i as above. Lifting f_1 to $f_1 \circ \pi \in L^\infty(\tilde{\mu}_1)$, on this larger system we know that it can be approximated in $L^2(\tilde{\mu}_1)$ by finite sums of the form

$$\sum_p g_p \cdot h_{2,p} \cdots h_{k,p},$$

where $g_p \in L^\infty(\tilde{\mu}_1)$ is invariant under $\langle \text{img } \tilde{\varphi}_1 \rangle$ and each $h_{i,p} \in L^\infty(\tilde{\mu}_i)$ is invariant under $\langle \text{img } \tilde{\varphi}_1 \tilde{\varphi}_i^{-1} \rangle$.

Appealing first to the uniform continuity of the operators $A_T^{\tilde{\lambda}}$ in each entry separately, and then to the linearity of these operators in the first entry, it therefore suffices to prove convergence of the averages

$$A_T^{\tilde{\lambda}}(f_1, \dots, f_k)$$

whenever f_1 is one such product function. However, this case lands within the hypothesis of the preceding lemma, which converts these into averages of the form

$$E(g \cdot A_T^\theta(1, h_2(f_2 \circ \eta_2), \dots, h_k(f_k \circ \eta_k)) \mid \eta_0).$$

The norm convergence of these now follows from the norm convergence of the averages $A_T^\theta(1, h_2(f_2 \circ \eta_2), \dots, h_k(f_k \circ \eta_k))$, which is promised by the inductive hypothesis applied to the simpler polynomial family $(\varphi_2, \dots, \varphi_k)$.

Property (2) Of course, property (1) already implies convergence of the aver-

aged couplings

$$\int_0^T (\text{id}_{X_0} \times u_1^{\varphi_1(t,h)} \times u_2^{\varphi_2(t,h)} \times \cdots \times u_k^{\varphi_k(t,h)})_* \lambda \, dt$$

as $T \rightarrow \infty$ to some limit λ^h . We must next show that for any tuple of functions $f_i \in L^\infty(\mu_i)$, the λ^h -integrals are the same whether we integrate $f_0 \otimes f_1 \otimes \cdots \otimes f_k$ or $(f_0 \circ u^{g_0}) \otimes (f_1 \circ u^{g_1}) \otimes \cdots \otimes (f_k \circ u^{g_k})$ for any

$$(g_0, g_1, \dots, g_k) \in G^{\Delta(k+1)} \quad \text{or} \quad (g_0, g_1, \dots, g_k) \in \langle \text{img } \vec{\varphi}(\cdot, h) \rangle.$$

This will give the invariance of λ^h under the u_\times action of $\langle G^{\Delta(k+1)} \cup \text{img } \vec{\varphi}(\cdot, h) \rangle$.

As in the case of property (1), in this step we can fix a choice of h and replace G with the subgroup $G^h := \langle \text{img } \varphi_1(\cdot, h) \cup \cdots \cup \text{img } \varphi_k(\cdot, h) \rangle$ if necessary, so that we may assume $r = 0$.

Since

$$E(f_1 \mid C\Sigma_1) \circ u_1^g = E(f_1 \circ u_1^g \mid C\Sigma_1)$$

for any g , by Proposition 8.2 it again suffices to treat the case when f_1 is $(C\Sigma_1)$ -measurable. Now we may consider again the previous construction of the \tilde{G} -systems $\tilde{\mathbf{X}}_i$ and \mathbf{Y}_i and their joinings λ' and θ . In these terms we wish to prove that

$$\begin{aligned} \int_{\prod_i \tilde{X}_i} f_0 \otimes f_1 \otimes \cdots \otimes f_k \, d\tilde{\lambda}' \\ = \int_{\prod_i \tilde{X}_i} (f_0 \circ \tilde{u}_0^{\tilde{g}_0}) \otimes (f_1 \circ \tilde{u}_1^{\tilde{g}_1}) \otimes \cdots \otimes (f_k \circ \tilde{u}_k^{\tilde{g}_k}) \, d\tilde{\lambda}' \end{aligned}$$

for any tuple $f_i \in L^\infty(\tilde{\mu}_i)$ and any

$$(\tilde{g}_0, \tilde{g}_1, \dots, \tilde{g}_k) \in \tilde{G}^{\Delta(k+1)} \quad \text{or} \quad (\tilde{g}_0, \tilde{g}_1, \dots, \tilde{g}_k) \in \langle \text{img } \vec{\tilde{\varphi}}(\cdot) \rangle,$$

where $\tilde{\lambda}'$ is the limit joining obtained by averaging $\tilde{\lambda}$.

Arguing again as for property (1), by continuity and multilinearity we may now assume that f_1 is of the special form $g \cdot h_2 \cdots h_k$ assumed by Lemma 9.1, and so by that lemma it now suffices to prove that

$$\begin{aligned} \int_{Y_0 \times \tilde{X}_0 \times Y_2 \times \cdots \times Y_k} (g'(f_0 \circ \eta_0)) \otimes 1 \otimes \cdots \otimes (h'_k(f_k \circ \eta_k)) \, d\theta' \\ = \int_{Y_0 \times \tilde{X}_0 \times Y_2 \times \cdots \times Y_k} ((g'(f_0 \circ \eta_0)) \circ v_0^{\tilde{g}_0}) \otimes 1 \otimes \cdots \otimes ((h'_k(f_k \circ \eta_k)) \circ v_k^{\tilde{g}_k}) \, d\theta', \end{aligned}$$

where θ' is the limit joining obtained by averaging θ . With this re-arrangement the coordinate in \tilde{X}_1 vanishes from the picture, and what remains is just an instance of property (2) for the simpler tuple of polynomial maps $(\tilde{\varphi}_2, \dots, \tilde{\varphi}_k)$, which is known by induction.

Property (3) Lastly, we must show that there is a Zariski residual set $E \subseteq \mathbb{R}^r$ such that for any tuple of functions f_i the limit

$$\lim_{T \rightarrow \infty} \int_{X_0} f_0 \cdot A_T^\lambda(f_1, \dots, f_k) d\mu_0$$

is the same for all $h \in E$, which will imply that the map $h \mapsto \lambda^h$ is Zariski generically constant (and hence, by property (2), that this generic value must be invariant under the whole of $\langle G^{\Delta(k+1)} \cup \text{img } \vec{\varphi} \rangle$). In this step, of course, we may not restrict to a single value of h .

Clearly it suffices to prove this h -independence for functions f_i drawn from countable $\|\cdot\|_2$ -dense subsets of $L^\infty(\mu_i)$, and since a countable intersection of Zariski generic sets is Zariski generic we may therefore look for such a Zariski generic set for just a single tuple of functions f_i .

The full strength of Proposition 8.2 and our construction above now give a Zariski residual subset $E \subseteq \mathbb{R}^r$, extensions of \tilde{G} -systems $\pi : \tilde{\mathbf{X}}_i \rightarrow \mathbf{X}_i^{q(\cdot)}$ and a joining $\tilde{\lambda}$ of \tilde{G} -systems such that

$$\begin{aligned} & \int_{X_0 \times \dots \times X_k} f_0 \otimes \dots \otimes f_k d\lambda^h \\ &= \lim_{T \rightarrow \infty} \int_{\tilde{X}_0} (f_0 \circ \pi_0) \cdot A_T^{\tilde{\lambda}}(E(f_1 \circ \pi | \Lambda), f_2 \circ \pi_2, \dots, f_k \circ \pi_k) d\tilde{\mu}_0 \end{aligned}$$

for all $h \in E$, where now

$$\Lambda := \tilde{\Sigma}_1^{\langle \text{img } \tilde{\varphi}_1 \rangle} \vee \bigvee_{i=2}^k \tilde{\Sigma}_1^{\langle \text{img } \tilde{\varphi}_1 \tilde{\varphi}_i^{-1} \rangle}.$$

Clearly it suffices to show that the desired h -independence holds on some further Zariski residual subset of E , and now the same manipulations as above give a reduction of this to a proof that the limits

$$\lim_{T \rightarrow \infty} \int_{Y_0} (g'_0(f_0 \circ \eta_0)) \cdot A_T^\theta(1, h'_2(f_2 \circ \eta_2), \dots, h'_k(f_k \circ \eta_k)) d\nu_0$$

are independent of h on some Zariski residual set, where θ and the Y_i have been constructed from λ and the \tilde{X}_i as previously. The dependence on h in this expression is all in the off-diagonal polynomial trajectory that appears in the average

A_T^θ . Once again, the fact that this limit is generically constant now follows from the inductive hypothesis applied to the family $(\tilde{\varphi}_2, \dots, \tilde{\varphi}_k)$, and so the proof is complete. \square

10 Further questions

10.1 Other questions in continuous time

Theorems 1.1 and 1.2 suggest many possible extensions involving different kinds of averaging, just as for any other equidistribution phenomenon. The following paragraphs contain a sample of these possibilities.

First, given another connected nilpotent group G' , one could ask more generally about polynomial maps $\varphi_i : G' \rightarrow G$ and the resulting off-diagonal averages along a Følner sequence of subsets $F_N \subseteq G'$. Do these always converge as in our main theorems? This seems likely, and I suspect that the methods of proof above can provide significant insight into this question, but it may be tricky to set up the right generalization of PET induction.

A little more abstractly, the off-diagonal polynomial trajectory

$$\{(\varphi_1(t), \varphi_2(t), \dots, \varphi_k(t)) : t \in \mathbb{R}\}$$

is a semi-algebraic subset of G^k in the sense of real algebraic geometry (see, for instance, Bochnak, Coste and Roy [11]). Could it be that convergence as in Theorems 1.1 or 1.2 holds along the intersections of increasingly large balls with any semi-algebraic subset $V \subset G^k$, endowed with a suitable surface-area measure?

A more challenging question concerns the assumption that G be nilpotent. Do Theorems 1.1 or 1.2 still hold if we assume only that G is an arbitrary connected and simply connected Lie group? This is probably too much to ask, but some progress may be possible, for instance, if each φ_i has image lying within a unipotent subgroup of G . This seems a natural setting to investigate in view of Ratner's Theorems giving equidistribution and measure rigidity for unipotent flows on homogeneous spaces [34, 33, 35, 36], and Shah's extension of these results to averages over regular algebraic maps [37].

However, as remarked in the Introduction, the methods used to study homogeneous space flows are very different (and mostly much more delicate) from those explored in this paper. Shah's analysis of regular algebraic maps proceeds by first obtaining the invariance of a weak limit measure under some unipotent subgroup

and then using the resulting structure promised by Ratner's Theorems, whereas it is an essential feature of our inductive proof of Theorem 1.3 that the cases of homomorphisms φ_i and of more general polynomial maps must be treated together.

To illustrate more concretely some of the difficulties posed by non-nilpotent groups, consider the functional averages

$$\int_0^T (f_1 \circ u_1^t)(f_2 \circ u_2^t) dt$$

for a jointly measurable probability-preserving system (X, Σ, μ, u) for $G = \mathrm{SL}_2(\mathbb{R})$ and with $u_1, u_2 : \mathbb{R} \rightarrow \mathrm{SL}_2(\mathbb{R})$ parametrizing the upper- and lower-triangular subgroups respectively. (These averages are easily expressed in terms of the natural analog of Theorem 1.2.) If we assume that these averages do not tend to 0 for some choice of $f_1, f_2 \in L^\infty(\mu)$, then the van der Corput estimate and a rearrangement give also

$$\int_0^S \int_0^T \int_X f_1 \cdot (f_1 \circ u_1^s) \cdot ((f_2 \cdot (f_2 \circ u_2^s)) \circ u_2^t u_1^{-t}) d\mu dt ds \not\rightarrow 0$$

as $T \rightarrow \infty$ and then $S \rightarrow \infty$. In order to use this, we need some information about the averages along the trajectory $t \mapsto u_2^t u_1^{-t}$ in G . This is certainly a polynomial map in the sense of real algebraic geometry, but not in the sense of Definition 4.1, so further differencing does not seem to lead to a simplification of the problem. I have not examined in detail what other arguments (for example, using the representation theory of $\mathrm{SL}_2(\mathbb{R})$) might be brought to bear here, since this is only a very special case: it simply serves to illustrate that the method of PET induction cannot be applied so naively in this setting.

Finally, linked to the study of convergence and equidistribution is the problem of describing the limit joinings λ' . Some information on their possible structure is contained in the proof of Proposition 8.2 above, as remarked after that proposition, but it would be interesting to know whether they can be classified more precisely, possibly after extending each \mathbf{X}_i to a suitably-sated extension. A discussion of related issues in the setting of \mathbb{Z}^d -actions can be found in [1].

10.2 Discrete actions

Most past interest in the kind of off-diagonal average appearing in Theorem 1.1 has focused on actions of discrete groups. Suppose that Γ is a discrete nilpotent group, $\varphi_1, \varphi_2, \dots, \varphi_k : \mathbb{Z} \rightarrow \Gamma$ are polynomial maps (according to the obvious relative

of Definition 4.1), $\mathbf{X}_i = (X_i, \Sigma_i, \mu_i, T_i)$ are probability-preserving Γ -systems for $1 \leq i \leq k$ and λ is a joining of the systems \mathbf{X}_i . Much recent work has been directed towards understanding whether the off-diagonal averages

$$\frac{1}{N} \sum_{n=1}^N (T_1^{\varphi_1(n)} \times \cdots \times T_k^{\varphi_k(n)})_* \lambda$$

converge to some limit joining as $N \rightarrow \infty$, or whether the associated functional averages converge. Several partial results have appeared, and at the time of this writing Miguel Walsh has just settled the general case in his preprint [43].

Walsh's approach does not use heavy ergodic-theoretic machinery. It relies on reformulating the problem of norm convergence for the functional averages into a problem asking for some 'quantitative' guarantee that one can find long intervals of times N in which those averages are all close in $\|\cdot\|_2$. This new assertion can then be proved by a clever induction on the tuple of polynomial maps $(\varphi_1, \dots, \varphi_k)$, which is apparently different from Bergelson's PET induction.

In making this reformulation, Walsh uses ideas that have some precedent in Tao's proof of convergence when $\Gamma = \mathbb{Z}^d$ and all the φ_i are linear ([40]). Some of these ideas lie outside more traditional ergodic-theoretic approaches to this class of questions (such as the present paper), and they have the consequence that very little can be gleaned about the structure of the limits (functions or joinings). Therefore it would still be of interest to see a proof that gives some additional information, similar to our Theorem 1.3 or to the earlier, even more precise results of [21] or [44] in the case of discrete powers of a single transformation. We finish with an informal discussion of the difficulties that face any attempt to adapt the arguments of the preceding sections to the setting of discrete Γ .

The first and most obvious difficulty is that if these averaged couplings do converge to some limit λ' , it need not be invariant under the off-diagonal subgroup

$$\langle \text{img}(\varphi_1, \dots, \varphi_k) \rangle \leq \Gamma^k.$$

Indeed, let $\Gamma = \mathbb{Z}$, let $\varphi_1 \equiv 0$ and $\varphi_2(n) := n^2$, and let $\mathbf{X}_1 = \mathbf{X}_2$ be the system given by the generator rotation on $\mathbb{Z}/4\mathbb{Z}$. Since all square numbers are congruent to either 0 or 1 mod 4, it is easily computed that the limit obtained by averaging the diagonal joining λ is simply

$$\frac{1}{2}\lambda + \frac{1}{2}(\text{id} \times T)_*\lambda,$$

which is not $(\text{id} \times T)$ -invariant.

Of course, this is a trivial example, but it is not clear whether this kind of arithmetic system, appearing as a factor of more general systems \mathbf{X}_i , is the only possible obstruction to the desired extra invariance of the limit joining.

While this example bears only on the possible symmetries of the limit joining, in the continuous-time setting those symmetries play a crucial rôle in the proof of Proposition 8.2 above, and so the whole method of proof we have used in this paper may need substantial modification before it can give convergence results in the discrete-time world.

A second difficulty worth remarking is the absence of any useful replacement for the notion of Zariski genericity in the discrete-time setting. Of course, Corollary 5.5 is still true for discrete group actions: the problem is that it tells us nothing, because these groups are themselves countable.

It might be worth exploring a more subtle appeal to the reasoning of Corollary 5.3 in place of Corollary 5.5. The statement of Corollary 5.3 is also still true for discrete groups provided the subgroups H_1 and H_2 are both normal in $\langle H_1 \cup H_2 \rangle$. One possibility might begin as follows. If $\mathfrak{H}_1, \mathfrak{H}_2, \dots$, is a sequence of closed subspaces of a Hilbert space \mathfrak{H} , any two of which are relatively orthogonal over some common further subspace \mathfrak{K} , and if in addition $x \in \mathfrak{H}$ is such that $\inf_n \|P_n x\| > 0$ with P_n the orthoprojection onto \mathfrak{H}_n , then x also has a nonzero projection onto \mathfrak{K} (for otherwise the $P_n x$ would be an infinite sequence of mutually orthogonal projections of a single vector, all of them large, contradicting Bessel's Inequality).

Structure like this has previously been identified within orthogonal representations of a finitely generated nilpotent group by Leibman [24]. Using this reasoning, for example, one can show that if

$$\Gamma = \langle a, b \mid [a, b] =: c \text{ is central} \rangle$$

is the discrete Heisenberg group and $T : \Gamma \curvearrowright (X, \Sigma, \mu)$ is any action of it, then the σ -subalgebras

$$\Sigma^{\langle a \rangle} := \{A \in \Sigma : \mu(T^a A \Delta A) = 0\}$$

and $\Sigma^{\langle b \rangle}$ are relatively independent over the fully invariant factor Σ^T , even though in this discrete setting it can happen that $\Sigma^{\langle a \rangle} \neq \Sigma^{\langle a \rangle^n}$ and $\Sigma^{\langle a \rangle}$ is not globally T -invariant. This follows because a judicious appeal to the discrete version of Corollary 5.3 implies that the σ -algebras

$$\Sigma^{b^k \langle a \rangle b^{-k}}, \quad k \in \mathbb{Z},$$

are all relatively independent over $\Sigma^{\langle a, c \rangle}$, where $\langle a, c \rangle$ is normal in Γ . If now f and g are T^a - and T^b -invariant respectively, then applying T^b gives

$$\int f \cdot \mathbb{E}(g \mid \Sigma^{\langle a \rangle}) d\mu = \int (f \cdot \mathbb{E}(g \mid \Sigma^{\langle a \rangle})) \circ T^{b^k} d\mu = \int (f \circ T^{b^k}) \cdot (\mathbb{E}(g \mid \Sigma^{b^{-k}\langle a \rangle b^k})) d\mu.$$

Therefore the non-vanishing of this integral implies that g actually has uniformly nonzero conditional expectation onto every $\Sigma^{b^k\langle a \rangle b^{-k}}$. Hence by the argument sketched above, it must actually have nonzero conditional expectation onto $\Sigma^{\langle a, c \rangle}$, and similarly f must have nonzero conditional expectation onto $\Sigma^{\langle b, c \rangle}$. These two σ -algebras are now globally T -invariant and relatively independent over Σ^T , so putting this together shows that $\Sigma^{\langle a \rangle}$ and $\Sigma^{\langle b \rangle}$ are themselves relatively independent over Σ^T .

In order to use a similar idea to study off-diagonal or multiple averages, one might, for instance, try to prove a discrete analog of Proposition 8.2 according to which the characteristic factors Λ_h obtained depending on h are not mostly equal to each other, but are all relatively orthogonal over some common smaller σ -algebra Λ' . Then it might be possible to replace Λ^h with Λ' in subsequent arguments and gain more purchase on the asymptotic behaviour of our averages as a result. However, I do not have a precise statement to formulate based on this speculation.

Acknowledgements This work was supported by a research fellowship from the Clay Mathematics Institute. Much of it was carried out during a visit to the Isaac Newton Institute for the Mathematical Sciences. \triangleleft

A A continuous-time van der Corput estimate

We recall here for completeness a continuous-time variant of the classical van der Corput estimate for bounded Hilbert-space-valued sequences. The discrete-time version can be found in Section 1 of [17], and a continuous-time version in Appendix B of Potts [31].

Lemma A.1 *If $u : [0, \infty) \rightarrow \mathfrak{H}$ is a bounded strongly measurable map into a Hilbert space, then vector-valued non-convergence*

$$\int_0^T u(t) dt \not\rightarrow 0 \quad \text{as } T \rightarrow \infty$$

implies the scalar-valued non-convergence

$$\int_0^S \int_0^T \langle u(t+s), u(t) \rangle dt ds \not\rightarrow 0 \quad \text{as } T \rightarrow \infty \text{ and then } S \rightarrow \infty.$$

□

References

- [1] T. Austin. Multiple recurrence and the structure of probability-preserving systems. Unpublished survey, available online at [arXiv.org](https://arxiv.org/abs/1006.0491): 1006.0491.
- [2] T. Austin. Norm convergence of continuous time polynomial multiple ergodic averages. To appear, *Ergodic Theory Dynam. Systems*.
- [3] T. Austin. Pleasant extensions retaining algebraic structure, I. Preprint, available online at [arXiv.org](https://arxiv.org/abs/0905.0518): 0905.0518.
- [4] T. Austin. Pleasant extensions retaining algebraic structure, II. Preprint, available online at [arXiv.org](https://arxiv.org/abs/0910.0907): 0910.0907.
- [5] T. Austin. On the norm convergence of nonconventional ergodic averages. *Ergodic Theory Dynam. Systems*, 30(2):321–338, 2009.
- [6] T. Austin. Deducing the multidimensional Szemerédi theorem from an infinitary removal lemma. *J. Anal. Math.*, 111:131–150, 2010.
- [7] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems*, 7(3):337–349, 1987.
- [8] V. Bergelson. Ergodic Ramsey Theory – an Update. In M. Pollicott and K. Schmidt, editors, *Ergodic Theory of \mathbb{Z}^d -actions: Proceedings of the Warwick Symposium 1993-4*, pages 1–61. Cambridge University Press, Cambridge, 1996.
- [9] V. Bergelson and A. Leibman. A nilpotent Roth theorem. *Invent. Math.*, 147(2):429–470, 2002.
- [10] V. Bergelson and A. Leibman. Failure of the Roth theorem for solvable groups of exponential growth. *Ergodic Theory Dynam. Systems*, 24(1):45–53, 2004.
- [11] J. Bochnak, M. Coste, and M.-F. Roy. *Real algebraic geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1998. Translated from the 1987 French original, revised by the authors.

- [12] J. Bourgain. Double recurrence and almost sure convergence. *J. Reine Angew. Math.*, 404:140–161, 1990.
- [13] Q. Chu, N. Frantzikinakis, and B. Host. Commuting averages with polynomial iterates of distinct degrees. Preprint, available online at [arXiv.org:0912.2641](https://arxiv.org/abs/0912.2641).
- [14] L. J. Corwin and F. P. Greenleaf. *Representations of Nilpotent Lie Groups and their Applications. Part I*, volume 18 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990. Basic theory and examples.
- [15] H. Furstenberg. Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, 31:204–256, 1977.
- [16] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi Theorem for commuting transformations. *J. d'Analyse Math.*, 34:275–291, 1978.
- [17] H. Furstenberg and B. Weiss. A mean ergodic theorem for $\frac{1}{N} \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$. In V. Bergleson, A. March, and J. Rosenblatt, editors, *Convergence in Ergodic Theory and Probability*, pages 193–227. De Gruyter, Berlin, 1996.
- [18] E. Glasner. *Ergodic Theory via Joinings*. American Mathematical Society, Providence, 2003.
- [19] B. Host. Ergodic seminorms for commuting transformations and applications. *Studia Math.*, 195(1):31–49, 2009.
- [20] B. Host and B. Kra. Convergence of polynomial ergodic averages. *Israel J. Math.*, 149:1–19, 2005. Probability in mathematics.
- [21] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. Math.*, 161(1):397–488, 2005.
- [22] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. Wiley-Interscience [John Wiley & Sons], New York, 1974. Pure and Applied Mathematics.
- [23] A. Leibman. Multiple recurrence theorem for measure preserving actions of a nilpotent group. *Geom. Funct. Anal.*, 8(5):853–931, 1998.
- [24] A. Leibman. The structure of unitary actions of finitely generated nilpotent groups. *Ergodic Theory Dynam. Systems*, 20(3):809–820, 2000.

- [25] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Israel J. Math.*, 146:303–315, 2005.
- [26] E. Lesigne, B. Rittaud, and T. de la Rue. Weak disjointness of measure-preserving dynamical systems. *Ergodic Theory Dynam. Systems*, 23(4):1173–1198, 2003.
- [27] G. A. Margulis. Dynamical and ergodic properties of subgroup actions on homogeneous spaces with applications to number theory. In *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*, pages 193–215, Tokyo, 1991. Math. Soc. Japan.
- [28] D. W. Morris. *Ratner’s theorems on unipotent flows*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2005.
- [29] I. B. S. Passi. Polynomial maps on groups. *J. Algebra*, 9:121–151, 1968.
- [30] I. B. S. Passi. *Group rings and their augmentation ideals*, volume 715 of *Lecture Notes in Mathematics*. Springer, Berlin, 1979.
- [31] A. Potts. Multiple ergodic averages for flows and an application. Preprint, available online at [arXiv.org: 0910.3687](https://arxiv.org/abs/0910.3687), 2009.
- [32] C. Pugh and M. Shub. Ergodic elements of ergodic actions. *Compositio Math.*, 23:115–122, 1971.
- [33] M. Ratner. On measure rigidity for unipotent subgroups of semisimple lie groups. *Acta Math.*, 165:229–309, 1990.
- [34] M. Ratner. Strict measure rigidity for unipotent subgroups of solvable groups. *Invent. Math.*, 101:449–482, 1990.
- [35] M. Ratner. On Raghunathan’s measure conjecture. *Ann. of Math. (2)*, 134(3):545–607, 1991.
- [36] M. Ratner. Raghunathan’s topological conjecture and distributions of unipotent flows. *Duke Math. J.*, 63(1):235–280, 1991.
- [37] N. A. Shah. Limit distributions of polynomial trajectories on homogeneous spaces. *Duke Math. J.*, 75(3):711–732, 1994.
- [38] A. N. Starkov. *Dynamical systems on homogeneous spaces*, volume 190 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2000. Translated from the 1999 Russian original by the author.

- [39] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.
- [40] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28:657–688, 2008.
- [41] T. Tao and V. Vu. *Additive combinatorics*. Cambridge University Press, Cambridge, 2006.
- [42] V. S. Varadarajan. *Geometry of quantum theory*. Springer-Verlag, New York, second edition, 1985.
- [43] M. N. Walsh. Norm convergence of nilpotent ergodic averages. To appear, *Ann. Math.*
- [44] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, 20(1):53–97 (electronic), 2007.

Mathematics Department, Brown University,
 Box 1917, 151 Thayer Street,
 Providence, RI 02912, USA
 timaustin@math.brown.edu,
 www.math.brown.edu/~timaustin